

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:22:25 ; Search time 97.3985 Seconds
(without alignments)
2272.479 Million cell updates/sec

Title: US-10-658-989A-2
Perfect score: 3488
Sequence: 1 GPGEPGPTGLPFPBGRGG.....GEOGVFDLGPFGSPGAG 617

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_23Sep04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3488	100.0	617	8 ADM48391	Adm48391 Human rec
2	3488	100.0	617	8 ADQ26217	Adq26217 Human gel
3	3482	99.8	821	8 ADM48392	Adm48392 Human rec
4	2237.5	64.1	1057	3 AAY84544	Aay84544 A human c
5	2237.5	64.1	1058	3 AAY84403	Aay84403 Amino aci
6	2237.5	64.1	1107	3 AAY84540	Aay84540 Amino aci
7	2237.5	64.1	1171	3 AAY84538	Aay84538 A chimeri
8	2237.5	64.1	1464	7 AAU68485	AAu68485 Human rec
9	2237.5	64.1	1464	7 ADD45059	Add45059 Human Pro
10	2237.5	64.1	1464	7 ADD45055	Add45055 Human Pro
11	2237.5	64.1	1464	7 ADD45051	Add45051 Human Pro
12	2234	64.0	1449	4 AAU02535	Aau02535 Porcine a
13	2233.5	64.0	1161	7 ADE87050	Ade87050 Human pan
14	2233.5	64.0	1461	5 AEG93947	Aeg93947 Human pol
15	2233.5	64.0	1464	4 AAU14136	Aau14136 Human nov
16	2233.5	64.0	1536	7 ADE87051	Ade87051 Human pan
17	2231.5	64.0	1057	3 AAY84541	Aay84541 Amino aci
18	2229.5	63.9	1388	3 AAY84539	Aay84539 Amino aci
19	2228.5	63.9	1341	2 AAR71701	Aar71701 Collagen
20	2228.5	63.9	1341	3 AAY96122	Aay96122 Collagen
21	2228.5	63.9	1341	5 AAU18475	Aau18475 Human col
22	2228.5	63.9	1341	5 ABB80733	Abb80733 Collagen
23	2228.5	63.9	1341	5 ABB09625	Abb09625 Amino aci
24	2228.5	63.9	1341	7 ADF13075	Adf13075 Human col
25	2227.5	63.9	1464	4 AAB82454	Aab82454 Human pro

26	2227.5	63.9	1464	5 ABB90764	Abb90764 Human Tum
27	2227.5	63.9	1464	5 ABP68610	Abp68610 Human pan
28	2227.5	63.9	1464	6 ABUS4471	Abus4471 Human Tum
29	2227.5	63.9	1464	6 ABR47417	AbR47417 Breast ca
30	2227.5	63.9	1464	6 ABR92064	AbR92064 Human cer
31	2227.5	63.9	1464	7 ADD14142	Add14142 Human src
32	2227.5	63.9	1464	7 ADP65246	Adp65246 Human alp
33	2227.5	63.9	1464	8 ADQ19470	Adq19470 Human sof
34	2225.5	63.8	1463	4 AAU02532	Aau02532 Bovine al
35	2224.5	63.8	1464	7 ADE87048	Ade87048 Human pan
36	2224.5	63.7	1107	2 AAR89472	Aar89472 Collagen/
37	2222.5	63.7	1169	3 AAR89469	Aar89469 Collagen/
38	2222.5	63.7	1169	3 AAY84537	Aay84537 Amino aci
39	2222.5	63.7	1171	2 AAR89470	Aar89470 Collagen/
40	2222.5	63.7	1388	2 AAR89471	Aar89471 Collagen/
41	2222	63.7	1211	7 ADE87057	Ade87057 Human pan
42	2222	63.7	1226	7 ADE87062	Ade87062 Human pan
43	2222	63.7	1411	3 AAY56800	Aay56800 Human pre
44	2215.5	63.5	1453	7 ADD45053	Add45053 Rat Prote
45	2215.5	63.5	1453	7 ADD45057	Add45057 Rat Prote

ALIGNMENTS

RESULT 1
ADM48391
ID ADM48391 standard; protein; 617 AA.
XX
AC ADM48391;
XX

03-JUN-2004 (first entry)

Human recombinant gelatin-like polypeptide Hu-3.

Plasma substitute; Gelatin-like protein; plasma expander; human.

Homo sapiens.

EP1398324-A1.

17-MAR-2004.

11-SEP-2002; 2002EP-00078745.

11-SEP-2002; 2002EP-00078745.

(FUJF) FUJI PHOTO FILM BV.

Bouwstra JB, Toda Y;

WPI; 2004-229415/22.

Composition useful as substitute for plasma, comprises solution of saline and recombinant gelatin-like protein having colloid osmotic function.

Example 1; SEQ ID NO 2; 3lpp; English.

The present sequence is the protein sequence of recombinant gelatin-like protein Hu-3. This is a trimer of human recombinant gelatin-like protein Hu-1 ADM48390. A claimed composition suitable as a substitute for plasma comprises a solution of saline and a protein having a colloid osmotic function. The protein is a recombinant gelatin-like protein with a molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The recombinant gelatin-like protein, or a dimer, trimer or tetramer of the protein, is useful as a plasma expander that has a lower clearance rate from blood circulation, provides better and predictable regulation of clearance rate and which is less susceptible to proteolytic degradation than presently used gelatin derivatives. Recombinant gelatin-like proteins that are in essence free of hydroxyproline do not give rise to an immunological reaction with blood samples containing Igg antibodies. The gelatin-like proteins can be covalently attached to pharmaceutically

CC active compounds. After administration, the coupled medicament will not
CC diffuse from the circulating blood into the interstitium. Clearance by
CC liver and kidney will be kept to a minimum, ensuring a more constant
CC plasma level of the medicament. Suitable medicaments include those
CC involved in intervening blood clotting, vasodilation, function of
CC erythrocytes, thrombocytes and leukocytes, immune responses, and blood
CC levels of messenger molecules such as hormones.

XX
SQ Sequence 617 AA;
Query Match 100.0%; Score 3488; DB 8; Length 617;
Best Local Similarity 100.0%; Pred. No. 1.7e-207; Mismatches 0; Indels 0; Gaps 0;
Matches 617; Conservative 0;
QY 1 GPPGEPGPTGLPFPGERGSGRGGFCADGVAGPKGAGERSGPGAGKSGSPGAGRP 60
Db 1 GPPGEPGPTGLPFPGERGSGRGGFCADGVAGPKGAGERSGPGAGKSGSPGAGRP 60
QY 61 GEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPPPGARGQAGVGMFPKGA 120
Db 61 GEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPPPGARGQAGVGMFPKGA 120
QY 121 GEPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 180
Db 121 GEPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 180
QY 181 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 240
Db 181 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 240
QY 241 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 300
Db 241 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 300
QY 301 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 360
Db 301 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 360
QY 361 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 420
Db 361 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 420
QY 421 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 480
Db 421 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 480
QY 481 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 540
Db 481 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 540
QY 541 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 600
Db 541 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 600
QY 601 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 617
Db 601 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 617

RESULT 2
ADQ26217
ID ADQ26217 standard; protein; 617 AA.
XX
XX
AC ADQ26217;
XX
DT 23-SEP-2004 (first entry)
XX
DE Human gelatine-like polypeptide Hu-3.
XX
KW Human; gelatine-like protein; Hu-3; microcarrier; cell culture.
XX
OS Homo sapiens.
XX

PN WO2004056976-A2.
XX
PD 08-JUL-2004.
XX
PF 23-DEC-2003; 2003WO-NL000922.
XX
PR 23-DEC-2002; 2002EP-00080539.
XX
PA (FUUF) FUJI PHOTO FILM BV.
XX
PI Bouwstra JB, Van Es AJJ, Toda Y;
XX
XX WPI; 2004-507711/48.
XX
PT Preparing cell culture support useful for culturing anchorage dependent
PT cells, involves coating microcarrier bead with gelatine or gelatine-like
PT protein having specific molecular weight.
XX
PS Example 1; SEQ ID NO 1; 19pp; English.
XX
CC The present sequence is the protein sequence of human recombinant gelatin
CC -like polypeptide Hu-3, which has a molecular weight of approximately 54
CC kDa and which contains multiple Gly-Xaa-Yaa triplets. In an example from
CC the invention, recombinant Hu-3 was immobilised on polystyrene beads
CC using a heterobifunctional crosslinking agent, and used in a spinner
CC flask cell culture. A claimed process for the preparation of a cell
CC culture support comprises the step of coating a microcarrier bead with
CC gelatine or gelatine-like protein having a molecular weight of about 40-
CC 200 kDa, and optionally further comprising the step of immobilising the
CC gelatine or gelatine-like protein on the microcarrier. In this process,
CC more than 75%, preferably more than 85% and more preferably more than 95%
CC of the gelatine or gelatine-like protein has the same molecular weight.
CC The gelatine or gelatine-like protein is recombinantly produced to obtain
CC a material of uniform molecular weight and to reduce the risk of
CC contamination with prions. A claimed cell support consists of microbeads
CC of 50-500 um size coated with a gelatine-like protein consisting of at
CC least 95% Gly-Xaa-Yaa triplets and containing at least 15% proline
CC residues and less than 5% of hydroxyproline residues, with a molecular
CC weight distribution showing a maximum between 40 and 200 kDa, at least
CC 75% of the protein molecules having a molecular weight within 2% of the
CC maximum. Large-scale production of expressed products can be accomplished
CC with gelatine-coated microcarriers.

XX
SQ Sequence 617 AA;
Query Match 100.0%; Score 3488; DB 8; Length 617;
Best Local Similarity 100.0%; Pred. No. 1.7e-207; Mismatches 0; Indels 0; Gaps 0;
Matches 617; Conservative 0;
QY 1 GPPGEPGPTGLPFPGERGSGRGGFCADGVAGPKGAGERSGPGAGKSGSPGAGRP 60
Db 1 GPPGEPGPTGLPFPGERGSGRGGFCADGVAGPKGAGERSGPGAGKSGSPGAGRP 60
QY 61 GEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPPPGARGQAGVGMFPKGA 120
Db 61 GEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPPPGARGQAGVGMFPKGA 120
QY 121 GEPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 180
Db 121 GEPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 180
QY 181 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 240
Db 181 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 240
QY 241 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 300
Db 241 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 300
QY 301 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 360
Db 301 GPPGKAGRGVPPGAVGAPKDGAGAGQPPGAPGAGERGEGOGPAGSFGQGLPGPA 360

QY 361 GPAGERGQAGSPGPGQGLPGAGPGEAGKPGEGQVPGDLGAPGSPGAGPFGPTGLP 420
Db 361 GPAGERGQAGSPGPGQGLPGAGPGEAGKPGEGQVPGDLGAPGSPGAGPFGPTGLP 420
QY 421 GPPGERGCGSRGFPFGADGVAGPKGAGPAGERSGPGAGPKGSPGAGRPGEAGLPKAGLGT 480
Db 421 GPPGERGCGSRGFPFGADGVAGPKGAGPAGERSGPGAGPKGSPGAGRPGEAGLPKAGLGT 480
QY 481 GSPGSPGDKTGPFGAGODGRPPGPPGAGSGQAGVWGFPFGPKGAAGEPGKAGRGVVP 540
Db 481 GSPGSPGDKTGPFGAGODGRPPGPPGAGSGQAGVWGFPFGPKGAAGEPGKAGRGVVP 540
QY 541 GPPGAVGAGKDGAGAGQPPGAGPAGERGERGQAGSPGPGQGLPGAGPFGAGKPGEQ 600
Db 541 GPPGAVGAGKDGAGAGQPPGAGPAGERGERGQAGSPGPGQGLPGAGPFGAGKPGEQ 600
QY 601 GVPGLGAPGSPGAGG 617
Db 601 GVPGLGAPGSPGAGG 617

RESULT 3

ADM48392
ID ADM48392 standard; protein; 821 AA.
XX AC
XX ADM48392;
XX DT 03-JUN-2004 (first entry)
XX DE Human recombinant gelatin-like polypeptide Hu-4.
XX KW Plasma substitute; Gelatin-like protein; plasma expander; human.
XX OS Homo sapiens.
XX PN EPI398324-A1.
XX PD 17-MAR-2004.
XX PF 11-SEP-2002; 2002EP-00078745.
XX PR 11-SEP-2002; 2002EP-00078745.
XX PA (FUJF) FUJI PHOTO FILM BV.
XX PI Bouwstra JB, Toda Y;
XX DR WPI; 2004-229415/22.
XX PT Composition useful as substitute for plasma, comprises solution of saline
PT and recombinant gelatin-like protein having colloid osmotic function.
XX FS Example 1; SEQ ID NO 3; 31pp; English.

XX The present sequence is the protein sequence of recombinant gelatin-like
CC protein Hu-3. This is a tetramer of human recombinant gelatin-like
CC protein Hu-1. ADM48390. A claimed composition suitable as a substitute for
CC plasma comprises a solution of saline and a protein having a colloid
CC osmotic function. The protein is a recombinant gelatin-like protein with
CC a molecular weight of at least 10 kDa and at most 50 kDa and an
CC isoelectric point of less than 8. It is especially Hu-1 or Hu-deam
CC ADM48393. The recombinant gelatin-like protein, or a dimer, trimer or
CC tetramer of the protein, is useful as a plasma expander that has a lower
CC clearance rate from blood circulation, provides better and predictable
CC regulation of clearance rate and which is less susceptible to proteolytic
CC degradation than presently used gelatin derivatives. Recombinant gelatin-
CC like proteins that are in essence free of hydroxyproline do not give rise
CC to an immunological reaction with blood samples containing IgE
CC antibodies. The gelatin-like proteins can be covalently attached to
CC pharmaceutically active compounds. After administration, the coupled
CC medicament will not diffuse from the circulating blood into the
CC interstitium. Clearance by liver and kidney will be kept to a minimum,
CC ensuring a more constant plasma level of the medicament. Suitable

CC medicaments include those involved in intervening blood clotting,
CC vasodilation, function of erythrocytes, thrombocytes and leukocytes,
CC immune responses, and blood levels of messenger molecules such as
CC hormones.

XX Sequence 821 AA;

Query Match 99.8%; Score 3482; DB 8; Length 821;

Best Local Similarity 100.0%; Pred. No. 4.9e-207;

Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPPGSPGPTGLPSPGSGRGGFPFGADGVAGPKGAGPAGERSGPGAGPKGSPGAGRPGEAGRP 60
Db 1 GPPGSPGPTGLPSPGSGRGGFPFGADGVAGPKGAGPAGERSGPGAGPKGSPGAGRPGEAGRP 60
QY 61 GEAGLPKAGLGTGSPGSGRGGFPFGADGVAGPKGAGPAGERSGPGAGPKGSPGAGRPGEAGRP 120
Db 61 GEAGLPKAGLGTGSPGSGRGGFPFGADGVAGPKGAGPAGERSGPGAGPKGSPGAGRPGEAGRP 120
QY 121 GPPGKAGRGVFPFGAGVAGPKAGKDGAGAGQPPGAGPAGERSGPGAGPKGSPGAGRPGEAGRP 180
Db 121 GPPGKAGRGVFPFGAGVAGPKAGKDGAGAGQPPGAGPAGERSGPGAGPKGSPGAGRPGEAGRP 180
QY 181 GPPGKAGRGVFPFGAGVAGPKAGKDGAGAGQPPGAGPAGERSGPGAGPKGSPGAGRPGEAGRP 240
Db 181 GPPGKAGRGVFPFGAGVAGPKAGKDGAGAGQPPGAGPAGERSGPGAGPKGSPGAGRPGEAGRP 240
QY 241 GPAGERSGPGAGPKGSPGAGRPGEAGLPKAGLGTGSPGSGRGGFPFGADGVAGPK 300
Db 241 GPAGERSGPGAGPKGSPGAGRPGEAGLPKAGLGTGSPGSGRGGFPFGADGVAGPK 300
QY 301 GPPGPPGARGQAGVWGFPFGKGAAGEPGKAGRGVFPFGAGVAGPKAGKDGAGAGQPPGPA 360
Db 301 GPPGPPGARGQAGVWGFPFGKGAAGEPGKAGRGVFPFGAGVAGPKAGKDGAGAGQPPGPA 360
QY 361 GPAGERGQAGSPGPGQGLPGAGPGEAGKPGEGQVPGDLGAPGSPGAGPFGPTGLP 420
Db 361 GPAGERGQAGSPGPGQGLPGAGPGEAGKPGEGQVPGDLGAPGSPGAGPFGPTGLP 420
QY 421 GPPGERGCGSRGFPFGADGVAGPKGAGPAGERSGPGAGPKGSPGAGRPGEAGLPKAGLGT 480
Db 421 GPPGERGCGSRGFPFGADGVAGPKGAGPAGERSGPGAGPKGSPGAGRPGEAGLPKAGLGT 480
QY 481 GSPGSPGDKTGPFGAGODGRPPGPPGAGSGQAGVWGFPFGPKGAAGEPGKAGRGVVP 540
Db 481 GSPGSPGDKTGPFGAGODGRPPGPPGAGSGQAGVWGFPFGPKGAAGEPGKAGRGVVP 540
QY 541 GPPGAVGAGKDGAGAGQPPGAGPAGERGERGQAGSPGPGQGLPGAGPFGAGKPGEQ 600
Db 541 GPPGAVGAGKDGAGAGQPPGAGPAGERGERGQAGSPGPGQGLPGAGPFGAGKPGEQ 600
QY 601 GVPGLGAPGSPGAGG 616
Db 601 GVPGLGAPGSPGAGG 616

RESULT 4

AY84544

ID AY84544 standard; protein; 1057 AA.

XX AC

XX AY84544;

XX DT 25-JUL-2000 (first entry)

XX DE A human collagen 1 (alpha1) protein helical region.

XX KW Extracellular matrix protein; self aggregation; hydroxylated proline;
KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX OS Homo sapiens.
XX BP992586-A2.

XX PD 12-APR-2000.
 XX PF 07-OCT-1999; 99EP-00119184.
 XX PR 09-OCT-1998; 98US-00169768.
 XX PA (USSU) US SURGICAL CORP.
 XX PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX DR WPI; 2000-259138/23.
 XX DR N-PSDB; AAA12503.
 XX PT Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX PS Example 10; Fig 39A-E; 260pp; English.
 XX CC The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents human collagen I (alpha1) helical region,
 CC which may be produced using the method of the invention
 XX SQ Sequence 1057 AA;
 Query Match 64.18; Score 2237.5; DB 3; Length 1057;
 Best Local Similarity 60.23; Pred. No. 3e-130;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;
 QY 1 GPP-----GFPGTGLFPPGGRGGRGFPFGADGVAGPKPAGRGSPGPA 48
 DB |||||
 QY 297 GPPGAGEGKRGARGEPFTGLFPPGGRGGRGFPFGADGVAGPKPAGRGSPGPA 356
 DB |||||
 QY 49 GPKSGPEAGRPAGLPGAKLTGSPGSPGPKTKTPPGACQDGRPGPPPGARGQA 108
 DB |||||
 QY 357 GPKSGPEAGRPAGLPGAKLTGSPGSPGPKTKTPPGACQDGRPGPPPGARGQA 416
 QY 109 GVMGFPKGAAGPPGKAGRGVFPFGAVGPKADGEAGAQCPGPPGAPGAGRGQGPA 168
 DB |||||
 QY 417 GVMGFPKGAAGPPGKAGRGVFPFGAVGPKADGEAGAQCPGPPGAPGAGRGQGPA 476
 QY 169 GSPGFQGLPGAPPPGAGKPGGVGPDILGAPGSPGAGE-----PGP---- 212
 DB |||||
 QY 477 GSPGFQGLPGAPPPGAGKPGGVGPDILGAPGSPGAGE-----PGP---- 536
 QY 213 -----TGLPFGPPGGRGGRGFPFGADGVAGPKPAGRGSPGPKGSP 258
 DB |||||
 QY 537 GANCAPNDGAKGDAGAPGAPGSGGAPGLOMFGERGAGLPGKGDGRDAGPKGADGSP 596
 QY 259 GE-----AGRPGEAGLPGAK-----GLTSGPSGSPGDKTKTPPGAGQDGRP 300
 DB |||||
 QY 597 GKDGVRGLTGPIGPPGAPGAPGDKGSPGSPGAPGTGARGAPDGRGPPGPPGAPGPP 656
 QY 301 GPPGPPGARGQ-----AGVMGFPKPKGNAGERGPKKAGRGVFPGPPGA---- 341
 DB |||||

DB 657 GADGQPGAKGEPGDAGAKGDAGPPGPGAGPAGPPIGNVGAPGAKGARGSPGPGATGFP 716
 QY 342 -----VGPAGKDXGAGAGQPPGAPG-----AGERGEQPGAGSPGQGLPGPAGPGEAGKP 393
 DB 717 GAAGRVGPPGPGSGNAGPPGPPGAPGKGGKGRGETGAPRGVGGPPGPPGAGKGGSP 776
 QY 394 GEQGVPGDLGAPGPPGSPGAGEPGPTGLPAPPGERGGPSRGGADGVAGPKPAGRGSP 453
 DB 777 GADGFAGAPGTPGPGQIAGQGVVGLPGQRCGERGFPGLPGSPGEPKQGPSGASGERGPP 836
 QY 454 GPAGP---KGSFGEAGRGEGAGLPGAKLTGSPGSPGPDGKTGTPPGAGQDGRGPPGPP 510
 DB 837 GPMGPPGLAGPPGSGREGAFGAEGRDPSGAKGDRGETGAPGPPGAPGAPGAPGV 896
 QY 511 GARGQGVVGPFGPKGAAGEPKKAGRGVPPGPGAVGAPGAKDGE-----AGAQ 558
 DB 897 GPAGSGDRGETGTPAGPAGPVGPGARGPAGPQGPGRDKGETGQGRGKIKGHRGFSGLQ 956
 QY 559 GPPGP-----AGPAGERGEQPGAGSP---GFQGLPGPAGPPGAGKPGEQGV 603
 DB 957 GPPGPPGSGRQPSGASGAPGPRGPPGSGAGPKDGLNGLPGIPGPPGRGTGDAGPV 1016
 QY 604 GDLGAPGSPGAPG 616
 DB 1017 GPPGPPGPPGPPG 1029
 RESULT 5
 ID AAY84403
 XX AAY84403 standard; protein; 1058 AA.
 AC AAY84403;
 XX 12-JUL-2000 (first entry)
 DE Amino acid sequence of human type 1 (alpha1) collagen polypeptide.
 KW Alpha1 collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
 KW collagen; mussel adhesive protein; bioadhesive.
 XX Homo sapiens.
 XX WO200014201-A1.
 PD 16-MAR-2000.
 PF 07-SEP-1999; 99WO-US020462.
 PR 09-SEP-1998; 98US-0099652P.
 PA (USSU) US SURGICAL CORP.
 PA (PAOL) PAOLELLA D N.
 PA (GRUS) GRUSKIN E A.
 PA (BUEC) BUECHTER D D.
 XX Paolletta DN, Gruskin EA, Buechter DD;
 XX WPI; 2000-271051/23.
 DR N-PSDB; AAZ99843.
 XX Incorporating non-natural amino acid into polypeptide, useful e.g. for
 PT production of bioadhesives, by epoxidation or substitution of
 PT dehydroproline residues.
 XX Disclosure; Fig 6; 66pp; English.
 CC The present sequence represents a human type 1 (alpha1) collagen protein.
 CC Peptides derived from the protein were used to demonstrate incorporation
 CC of 3,4-dehydro-L-proline into the peptide, using the method of the
 CC invention. The specification describes a method for the incorporation of
 CC non-natural amino acid into a polypeptide. The method comprises reacting
 CC at least one 3,4-dehydroproline residue in the polypeptide with an
 CC epoxidation reagent from a polypeptide containing at least one 3,4-

CC epoxyproline residues. The method is used for studying the effects of non-
CC natural amino acids on structure and function of polypeptides. The method
CC is also useful for commercial production of collagen or mussel adhesive
CC proteins (which are useful as bioadhesives), and for incorporating a wide
CC variety of groups, including therapeutic ligands and biological probes,
CC into polypeptides.
XX
XX Sequence 1058 AA;
SQ

Query Match	54.1%	Score	2237.5	DB 3	Length	1058			
Best Local Similarity	60.2%	Pred.	No. 3e-130						
Matches	441	Conservative	23	Mismatches	152	Indels	117	Gaps	12
QY	1	GGP-----	-----CEPGTGLPGPPGRRGGSGRRFFGADGVAGPKGPA	AGPAGKGPAGERS	SGPGA	48			
DB	298	GGPPGAGERGKRGAR	GEPTGLPGPPGRRGGSGRFFGADGVAGPKGPA	AGPAGKGPAGERS	SGPGA	357			
QY	49	GPXGSPGEAGRPEAE	AGLPOAKGLTGSFSGPPDPGKTFPGPAGQDGRPGPPGARGQA	108					
DB	358	GPKGSPEAGRPEAE	GLPPOAKGLTGSFSGPPDPGKTFPGPAGQDGRPGPPGARGQA	417					
QY	109	GVWGPPGPKGAEPCK	AGKACERGVPVPGAVGVPAGKDGEGACAGCPGPPAGPAGERGQ	GPGA	168				
DB	418	GVWGFFGPKGAEPCK	AGKACERGVPVPGAVGVPAGKDGEGACAGCPGPPAGPAGERGQ	GPGA	477				
QY	169	GSPGFQGLPGPAP	GPGEAKGPEQGVVDLGA	PPSGPAGE-----	PGP-----	212			
DB	478	GSFGFQGLPGPAP	GPGEAKGPEQGVVDLGA	PPSGPAGE-----	PGP-----	537			
QY	213	-----	TGLPPGPRGRRGSGRRFFGADGVAGPKGPA	AGERS	SGPPAGK	GPSP	258		
DB	538	GANGAPNDGAKGD	AGAPAGPSQAGPGLQGMPEGERCAAGLP	PKGDRDAGPKG	ADGSG	597			
QY	259	GE-----	AGRPEAEGLPGAK-----	GLTSGSGSPGDPDKT	GPPGAGQ	GRP	300		
DB	598	GKDGVGLTGP	LOPPGAPAGPKDGSGSGSPGAPGTGARGADGRGEP	PPGPPAGPAGP	657				
QY	301	GPFPGPGARGQ-----	-----AGVMGFFGPKGAAGEPKAGERG	VGP	PPGA-----	341			
DB	658	GADGQFGAGXEP	GDAGAKGDGPPGPPAGPAGPPCPIGNVGAPKAGARGSAG	PPGATGFP	717				
QY	342	-----	VGPAGKDGEGAGAQDPPGAPG-----	AGERGEGPAGSPG	FOGLPGPAG	PPGAGKP	393		
DB	718	GAAGRYVGP	PPSGNAGPPGPPGAGXGKGPRGETGPARPGEVPPG	PPGPPGAGKGP	777				
QY	394	GEQGVPGDLGAP	SPGAPGPPGTGLPGPPGRRGGSGRRFFGADGVAGPKGPA	GERGSP	453				
DB	778	GADGPAGAP	PTPFGQTIAGRGVVVLPGQRRGSGFFCLPGPSGPKQK	QKQPSGASGER	837				
QY	454	GPAGP---	KQSPGEAGRPGEAGLPGAKGLTGS	SGSGPDGKTGPPGAGQDGR	PPGPP	510			
DB	838	GPWGPPCLAG	PPGSGESREGAPAGGSGPSGRDGS	PGAKGDRGETGAGPPGAPGAPG	PGPV	897			
QY	511	GARGQAGVMFP	PGPKNAGFPKGAERGVPVPGAVGPA	GKDG-----	AGAQ	558			
DB	898	GPAGKSGDRGET	GPAGPAGVPGAPAGKGPAGPQGRD	KGETGEGD	DRGTLKHUR	SGLQ	957		
QY	559	GPFGP-----	-----AGPAGRGEQ	QGAPSGP-----	GFOGLPGPAG	PPGAGKPG	QGV	603	
DB	958	GPFGPGSG	PEGQPSGASGAGPRGPPGSGAGATGK	CDGLNGLPIGPPGPRG	TGADGPV	1017			
QY	604	GDLAGPPSGPAG	616						
DB	1018	GPFGPPGPPGPP	1030						

RESULT 6
 AAY84540
 ID AAY84540 standard; protein; 1107 AA.
 XX
 AC AC AAY84540;
 XX
 DT 25-JUN-2000 (first entry)

XX	Amino acid sequence of a chimeric collagen 1 (alpha1)/decorin protein.
DE	
XX	Extracellular matrix protein; self aggregation, hydroxylated proline;
DE	trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW	collagen; fibrinogen; fibronectin; post translational hydroxylation;
KW	decorin; chimera.
XX	
OS	Homo sapiens.
OS	Unidentified.
OS	Chimeric.
XX	
XX	Key Location/Qualifiers
FH	Misc-difference 858
FT	/note= "Gly encoded by GCT"
FT	
XX	EP992586-A2.
PIN	
XX	12-APR-2000.
PD	
PP	07-OCT-1999; 99EP-00119184.
PF	
PR	09-OCT-1998; 98US-00169768.
XX	(USSU) US SURGICAL CORP.
XX	
XX	Gruskin EA, Buechter DD, Zhang G, Connolly K;
DR	WPI; 2000-259138/23.
DR	N-P5DB; AAA12500.
XX	
PT	Production of extracellular matrix proteins containing 4-trans-
PT	hydroxyproline results in native self aggregating proteins, useful on
PT	medical implants.
PS	
PS	Claim 24; Fig 18; 260pp; English.
XX	
CC	The specification describes a method for producing an extracellular
CC	matrix protein or its fragment. The extracellular matrix protein is
CC	capable of self aggregating in a cell which does not ordinarily
CC	hydroxylated prolines. The method comprises optimising a nucleic acid
CC	sequence for expression in the cell by substitution of codons preferred
CC	by that cell for naturally occurring codons not preferred by the cell;
CC	incorporating the nucleic acid sequence into the cell; and contacting the
CC	cell with a hypertonc growth medium containing at least one amino acid,
CC	selected from the group consisting of trans-4-hydroxyproline and 3-
CC	hydroxyproline to allow at least one of the amino acids to be assimilated
CC	into the cell and incorporated into the extracellular matrix protein. The
CC	method may be used to make host cells assimilate and incorporate trans-4-
CC	hydroxyproline into proteins. This is especially useful in the
CC	recombinant production of proteins such as collagen, fibrinogen and
CC	fibronectin whose ability to self aggregate and produce functional
CC	proteins depends on the post translational hydroxylation of proline. The
CC	method is also useful in studying the structure and function of
CC	polypeptides which do not normally contain trans-4-hydroxyproline. The
CC	present sequence represents a chimeric collagen 1 (alpha1)/decorin
CC	protein, which may be produced using the method of the invention
XX	
SQ	Sequence 1107 AA;

Query Match	64.1%;	Score	2237.5;	DB	3;	Length	1107;
Best Local Similarity	60.2%;	Prod. No.	3.1e-130;				
Matches	441;	Conservative	23;	Mismatches	152;	Indels	117;
Gaps	12;						

QY	1	GPP-----	GPQGTGLPGPGRGGPSRGGADGVAGPKPGAGERGSPGA	48
DB	297	GPPGAGEEKGARGEPGTLPGPPGRGGPSRGGADGVAGPKPGAGERGSPGA	356	
QY	49	GPKSPEAGRPAGLFGAKGLTSGPSGPDGKTGPPGPAQDGRGPPGPPGARGQA	108	
DB	357	GPKSPEAGRPAGLFGAKGLTSGPSGPDGKTGPPGPAQDGRGPPGPPGARGQA	416	
QY	109	GWVGFPKGAAGFPKGABRGVPPGPAVGPPAKGDGEAGAGQPGPGDGPAGERGEQQPA	168	

```

Db 417 GVMGFPKGAAGEPKAGRGVPPGAVGPAKDGGAQGGPPGAPAGERGEGGPA 476
QY 169 GSPGQGLPGAGPGEAGKGEQGVGDLGAPGSPGAGE-----PGP----- 212
Db 477 GSPGQGLPGAGPGEAGKGEQGVGDLGAPGSGARGGPPGGERGVQGGPPGAPGR 536
QY 213 -----TGLPSPGERGSGRFRFPAGDVGAGPKGAPAGERGSGPPAGPKGSP 258
Db 537 GANGAPNDGAKGADAGAPGSGQAGLQMPGERGAAGLPGPKGDRGDAGPKGADGSP 536
QY 259 GE-----AGRPGEAGLPGCAK-----GLTSGPSGPDGKTGPPGAGQDGRP 300
Db 597 KDGVRGLTGPIGPPGAPAGDKGESGSPGAPGFTGARGAPGDRGEPFPPGAPGAPPP 656
QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGEPKKAGRGVPPGPA----- 341
Db 657 GADQGGKAGGPPGDAGAKGDAGPPCPAGPAGPPGPIGNVAPGAKGARGSGAPPGATGFP 716
QY 342 -----VGPAGKDGAGAGQPPGAPG---AGERGEQGPAGSPGFGQLGPPAGPGEAGKP 393
Db 717 GAAGRVGPPGSGNAGPPGPPCPAGKGGKPRGETGPRGPPGPPGPPGAGEKGS 776
QY 394 GEQGVGDLGAPGSGPAGEPGTLGPPPPGERGSGRFPFGADGVAGPKGAPAGERGSP 453
Db 777 GADGPAGAPGTPGQGIAGQGVVGLPQGRGERGFPGLPGSPGEPKQKQSGASGERGPP 836
QY 454 GPAGP---KSGSPGAGRPCEAGLPGAKLTGSPGSPGPDGKTGPPGAGQDGRGPPGPP 510
Db 837 GPMGPPGLAGPPGSGREGAPGARGSPDRGSPGAKGDRGETGAPGPPGAPGAPGPPV 896
QY 511 GARGQAGVMGPPGPKGAAGEPKAGRGVPPGPAVGPAGKDG-----AGAQ 558
Db 897 GPAGKSGDRGTGPAGPAGPVGPPAGARGPAGPQGPGRGDKETGEGQGDGKIGHRGFSLQ 956
QY 559 GPPGP-----AGPAGERGEGGAPGSP---GFOGLPGPAGPPGAGRGVGGQVP 603
Db 957 GPPGPPGSGQSGASGAPRGPPGSGAGAPKGLNGLPGPIGPPGPRGRTGDGPV 1016
QY 604 GDLGAPGSPGAP 616
Db 1017 GPPGPPGPPGPPG 1029

```

RESULT 7
 ID AAY84538
 AC AAY84538
 AC AAY84538;
 DT 25-JUL-2000 (first entry)

A chimeric collagen 1 (alpha1)/TGF-beta1 protein.

Extracellular matrix protein; self aggregation; hydroxylated proline;
 trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 collagen; fibrinogen; fibronectin; post translational hydroxylation;
 ss. transforming growth factor-beta1; TGF-beta1; chimera.

Homo sapiens.
 Unidentified.
 Chimeric.

Key Location/Qualifiers
 Misc-difference 858 /note= "Gly encoded by GCT"

EP992586-A2.
 12-APR-2000.
 07-OCT-1999; 99EP-00119184.

```

PR 09-OCT-1998; 98US-00169768.
XX (USSU ) US SURGICAL CORP.
PA Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX WPI: 2000-259138/23.
XX N-PSDB; AAL12498.
PT Production of extracellular matrix proteins containing 4-trans-
PT hydroxyproline results in native self aggregating proteins, useful on
PT medical implants.
XX Claim 23; Fig 15; 260pp; English.
XX The specification describes a method for producing an extracellular
XX matrix protein or its fragment. The extracellular matrix protein is
XX capable of self aggregating in a cell which does not ordinarily
XX hydroxylated prolines. The method comprises optimising a nucleic acid
XX sequence for expression in the cell by substitution of codons preferred
XX by that cell for naturally occurring codons not preferred by the cell;
XX incorporating the nucleic acid sequence into the cell; and contacting the
XX cell with a hypertonic growth medium containing at least one amino acid,
XX selected from the group consisting of trans-4-hydroxyproline and 3-
XX hydroxyproline to allow at least one of the amino acids to be assimilated
XX into the cell and incorporated into the extracellular matrix protein. The
XX method may be used to make host cells assimilate and incorporate trans-4-
XX hydroxyproline into proteins. This is especially useful in the
XX recombinant production of proteins such as collagen, fibrinogen and
XX fibronectin whose ability to self aggregate and produce functional
XX proteins depends on the post translational hydroxylation of proline. The
XX method is also useful in studying the structure and function of
XX polypeptides which do not normally contain trans-4-hydroxyproline. The
XX present sequence represents chimeric collagen 1 (alpha1)/transforming the
XX growth factor-beta1 (TGF-beta1) protein, which may be produced using the
XX method of the invention
XX Sequence 1171 AA;

```

Query Match 64.1%; Score 2237.5; DB 3; Length 1171;
 Best Local Similarity 60.2%; Pred. No. 3.3e-130;
 Matches 441; Conservative 23; Mismatches 132; Indels 117; Gaps 12;

```

QY 1 GPP-----GEPGPTGLPFPGERGSGRFPFGADGVAGPKGAPAGERGSPGA 48
Db 297 GPPGPAEGEKRGARGEPGPTGLPFPGERGSGRFPFGADGVAGPKGAPAGERGSPGA 356
QY 49 GPKSGPGEAGPGEAGLPGAKLTGSPGSPGPDGKTGPPGAGQDGRGPPGPPGARGQA 108
Db 357 GPKSGPGEAGPGEAGLPGAKLTGSPGSPGPDGKTGPPGAGQDGRGPPGPPGARGQA 416
QY 109 GVMGFPKGAAGEPKKAGRGVPPGPAVGPAGKDGGAQGGPPGAPAGERGEGGPA 168
Db 417 GVMGFPKGAAGEPKKAGRGVPPGPAVGPAGKDGGAQGGPPGAPAGERGEGGPA 476
QY 169 GSPGQGLPGAGPGEAGKGEQGVGDLGAPGSPGAGE-----PGP----- 212
Db 477 GSPGQGLPGAGPGEAGKGEQGVGDLGAPGSGARGGPPGGERGVQGGPPGAPGR 536
QY 213 -----TGLPSPGERGSGRFRFPAGDVGAGPKGAPAGERGSGPPAGPKGSP 258
Db 537 GANGAPNDGAKGADAGAPGSGQAGLQMPGERGAAGLPGPKGDRGDAGPKGADGSP 596
QY 259 GE-----AGRPGEAGLPGCAK-----GLTSGPSGPDGKTGPPGAGQDGRP 300
Db 597 KDGVRGLTGPIGPPGAPAGDKGESGSPGAPGFTGARGAPGDRGEPFPPGAPGAPPP 656
QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGEPKKAGRGVPPGPA----- 341
Db 657 GADQGGKAGGPPGDAGAKGDAGPPCPAGPAGPPGPIGNVAPGAKGARGSGAPPGATGFP 716
QY 342 -----VGPAGKDGAGAGQPPGAPG---AGERGEQGPAGSPGFGQLGPPAGPGEAGKP 793

```

Db 717 GAAGRVGPPGSPGNAGPPGPPGAGKGGKGRGTGTPAGRPGEVGPDPGPPGAGKGGSP 776
 Qy 394 GEOGVPGDLGAPGSPGAGPGTGLPFPGERGSPGSRGFPAGDVGAGPKGPPAGRGSP 453
 Db 777 GADGPAGAPGTGPGQGIAGQGVVGLPGQGRGSGFPGLPGSPGPKGQKSPGASGERGPP 836
 Qy 454 GPAGP---KSPGAGPPGAGLPGAKGLTSGSPGPDGKTTPPPGAGDGRPPGPPGP 510
 Db 837 GPMGPPGLAGPPGSGREGAPGAGSGPRDGSFGAKGDRGTGPPGPPGAPGAPGPPV 996
 Qy 511 GARGQAGVMGPPGPKGAAGPGKAGRGVFPFGVAGVPAGKDG------AGAQ 558
 Db 897 GPAGKSGDRGTGTPAGPAGVPVGPAGARGPAGPQPRGDKGTGEGQGRGKIHGRGSGGLQ 956
 Qy 559 GPPGP-----AGPAGERGEGPPAGSP---GFQGLPAPGPPGAGPKGPPGQGV 603
 Db 957 GPPGPPGSPGQSPGASGAPGPRGPPGSGAGPKDGLNLPDIPGPPGPRGTGDPGV 1016
 Qy 604 GLGAPGSPGAG 616
 Db 1017 GPPGPPGPPGPPG 1029
 RESULT 8
 AAW68485
 ID AAW68485 standard; protein; 1464 AA.
 AC AAW68485;
 XX
 DT 08-DEC-1998 (first entry)
 XX
 DE Human recombinant collagen protein.
 XX
 KW Primer; PCR; amplification; human; collagen; mammal; plant; prosthesis;
 KW cardiac valve; ligament; tendon; skin; gingival implant; perfumes;
 KW nerve regeneration; antibiotic; growth factor; cancer; inflammatory;
 KW gelatin; glue; food.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 Key Location/Qualifiers
 FH Peptide 1..22
 FT /note= "signal peptide"
 FT 23..999
 FT /note= "mature protein"
 FT 161
 FT /note= "cleavage site for aminopeptidase"
 FT 1218
 FT /note= "cleavage site for carboxypeptidase"
 XX
 PN W09827202-A1.
 XX
 PD 25-JUN-1998.
 XX
 PF 17-DEC-1997; 97WO-FR002331.
 XX
 PR 17-DEC-1996; 96FR-00016224.
 XX
 PA (BIOC-) BIOCEM SA.
 XX
 PI Gruber V, Exposito J, Ruggiero F, Comte J, Garrone R, Merot B;
 PI Bournat P;
 XX
 DR WPI; 1998-362771/31.
 DR N-PSDB; AAV60814.
 XX
 XX New recombinant nucleic acid for expressing collagen or derivatives in
 PT plants - useful as, e.g. bio-materials and in therapeutic, cosmetic and
 PT odontological compositions.
 XX
 PS Disclosure; Fig 7; 138pp; French.
 XX

CC The invention relates to the production of mammalian collagen in plants.
 CC 2 clones: alpha3 and alpha22, spanning the human collagen type I gene
 CC were isolated from a MG-63 osteosarcoma library. Clone alpha3 contained
 CC 83 bp of the 5' untranslated region and the first 1920 bp of coding
 CC sequence, whereas clone alpha22 contained sequence encoding amino acids
 CC 171-1454 of the protein and around 500 bp of the 3' untranslated region.
 CC The 2 clones were used to generate a number of fragments which were used
 CC to construct composite sequences encoding variant collagen molecules. The
 CC fragments are: (A) containing nucleotides (nt) -4 to 479; (B) containing
 CC TAA upstream of the sequence encoding the PHS (pathogenesis-related
 CC protein S) signal peptide and bases 66-77 from the sequence encoding the
 CC N-terminus of the pro-collagen amino propeptide domain; (C) the whole of
 CC the amino propeptide domain (nt 72-479); (D) all of the amino-telopeptide
 CC domain (nt 474-534) and the N-terminus of the helical region (nt 535-
 CC 1920); (E) the DraIII-BamHI fragment (1709-2808) of alpha22, encoding aa
 CC 567-936 of the central helical domain; (F) the BamHI-EcoRI (2803-4362)
 CC region of alpha22, encoding aa 936-1192 in the central helical domain of the C-
 CC and aa 1193-1454 in the C-propeptide domain; (G) the C-terminus of the C-
 CC propeptide domain (aa 1346-1464) plus stop codons, and (H) as G but
 CC encoding aa 1343-1401 and also including the KDEL motif for retention in
 CC the ER. This sequence represents a recombinant human collagen. The
 CC encoding gene was constructed from fragments (A), (D), (E), (F) and (G).
 CC The recombinant gene is used for expression of mammalian collagen in
 CC plant cells. The transformed plants, their extracts and parts are useful
 CC as biomaterials (haemostatic compresses, sponges or bandages) and in
 CC pharmaceutical, medical, odontological, cosmetic and biotechnological
 CC compositions (e.g. as prostheses for cardiac valves, ligaments or tendons
 CC ; skin substitutes; gingival implants; microcapsules for perfumes; Guide
 CC tubes for nerve regeneration; slow release products for antibiotics;
 CC growth factors, anticancer agents or anti-inflammatory; surgical thread
 CC and components of ointments). They are suitable for treating any disorder
 CC related to collagen dysfunction and gelatin, produced from collagen, is
 CC used to produce glues, surgical prostheses and foods
 XX
 SQ Sequence 1464 AA;

Query Match 64.1%; Score 2237.5; DB 2; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 3.9e-130;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

Qy 1 GPP-----GEPGTGLPPPGRRGGSPGSRGFGDVGAGPKGPPAGRGSPGPA 48
 Db 458 GPPGPAEGEKGRCARGEPPTGLPPPGRRGGSPGSRGFGDVGAGPKGPPAGRGSPGPA 517
 Qy 49 GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDGKTTPPPGAGDGRPPGPPGARGOA 108
 Db 518 GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDGKTTPPPGAGDGRPPGPPGARGOA 577
 Qy 109 GVMGFPKXGAGEPKGAGERGVPPGAVGAGVAGDGEAGAGCGPPGPPAGRGSPGPA 168
 Db 578 GVMGFPKXGAGEPKGAGERGVPPGAVGAGVAGDGEAGAGCGPPGPPAGRGSPGPA 637
 Qy 169 GSPFGQLPFPAGPGEAGKPGEQGVFDLGAPGSPGAGE-----PGP---- 212
 Db 638 GSPFGQLPFPAGPGEAGKPGEQGVFDLGAPGSPGAGE-----PGP---- 212
 Qy 213 -----TGLPGRGRRGGSPGSRGFGDVGAGPKGPPAGRGSPGPA 258
 Db 698 GANGAPNDGAKDAGAPGSPGAPGQGLQMPGRRGAAGLPGPKDGRDAGPKGADGSP 757
 Qy 259 GE-----AGRPEAGLPGAK-----GLTSGSPGSPGDKTTPPPGAGDGRPP 300
 Db 758 GKDGVRGLTGPFGPPGAPAGDKGSGSPGAPGTGARGAPGDRGEPGPPGPPAGPAGPP 817
 Qy 301 GPPGPPGARGQ-----AGVMGFPKXGAGEPKGAGERGVPPGPA----- 341
 Db 818 GADGQPGKAGEPKDAGKGDAGPPGAPGAPGPPGTGNTVGAPGAKGARGSPGATGPP 877
 Qy 342 -----VGPAGKDEAGAQPPGPPAGP---AGERGEQGPAGSPGQGLPAGPPGAGKP 393
 Db 878 GAAGRVGPPGPPGSPGNAGPPGPPGAGKGGKGPREGTGPAGRPGEVGPDPGPPGAGKGP 937
 Qy 394 GEOGVPGDLGAPGSPGAPGPPGTGLPFPGERGSPGSRGFGDVGAGPKGPPAGRGSP 453

Db 938 GADGAPAGTPTGQIAGQWVGLPQGRGPFGLPGSGEPKQKQSGASGERGPP 997
 QY 454 GPAGP---KGSPEAGRPEAGLPCAKGLTSGSPGPDGKTGPPGAGQDGRPGPP 510
 Db 998 GPMGPPGLAGPPESGRGAPGAEPSGDRGSGPANGRGETGAGPPGAPGAPGPV 1057
 QY 511 GARGQAGVGMGPPKPAAGEPCKAGRGVPPGPPGAVGPAKDGK-----AGAQ 558
 Db 1058 GPAGKSGDRGTGTPAGPAGPVGPAGARGPAGPQGRGDKGTGQDRIKGRHGSGLQ 1117
 QY 559 GPPGP-----AGPAGEREQCPAGSP---GQGLPGAGPPGKAGKGEQGV 603
 Db 1118 GPPGPPGSGFSGSGSAGPAGPPGSGAGPAGKQGLNGLPGIPGPPGRGTGDGPV 1177
 QY 604 GDLGAPGSPGAP 616
 Db 1178 GPPGPPGPPGP 1190

RESULT 9

ADD45059
 ID ADD45059 standard, protein; 1464 AA.

AC ADD45059;

DT 29-JAN-2004 (first entry)

DE Human Protein P02452, SEQ ID NO 10491.

KW Human; pain; neuronal tissue; gene therapy;

KW spinal segmental nerve injury; chronic constriction injury; CCI;

KW spared nerve injury; SNL; Chung.

OS Homo sapiens.

PN WO2003016475-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-US025765.

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

PA (GENO) GEN HOSPITAL CORP.

PA (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

XX GENBANK; P02452.

FT New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a

CC method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the CC specification, a method for identifying a compound useful in treating CC pain and a pharmaceutical composition comprising the one or more CC polypeptides or their antibodies. The polynucleotide or the compound that CC modulates its activity is useful for preparing a medicament for treating CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction CC injury (CCI) and spared nerve injury (SNL)) in an animal (e.g. gene CC therapy). The sequence presented is a human protein (shown in Table 2 of CC the specification) which is differentially expressed during pain. Note: CC the sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic form directly from WIPO at CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1464 AA;

Query Match 64.1%; Score 2237.5; DB 7; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 3.9e-130;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPQPPGGRGPGSRGPPGADGVAGPKPAGERGSPGA 48
 Db 458 GPPGAGEGKGEARGEPCTGLPQPPGGRGPGSRGPPGADGVAGPKPAGERGSPGA 517
 QY 49 GPKGSPGAEGRPEAGLPCAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGQA 108
 Db 518 GPKGSPGAEGRPEAGLPCAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGQA 577
 QY 109 GVMGPPGPKGAAGEPCKAGRGVPPGPAAGVAGPKAGDGEAGAGQPPGPAAGEGEOGPA 168
 Db 578 GVMGPPGPKGAAGEPCKAGRGVPPGPAAGVAGPKAGDGEAGAGQPPGPAAGEGEOGPA 637
 QY 169 GSPGQGLPGPAGPPEAGKPGEQVPCDLGAPGSPGAGE-----PGP---- 212
 Db 638 GSPGQGLPGPAGPPEAGKPGEQVPCDLGAPGSPGARGRPFGERGVQPPGAPGR 697
 QY 213 -----TGLPQPPGGRGPGSRGPPGADGVAGPKPAGERGSPGAPGKSP 258
 Db 698 GANGAPGNDGAKGDAGAPGAGPQGLQMPGERGAAGLPGPKGDRGDAGPKGADGSP 757
 QY 259 GE-----AGRPEAGLPCAK-----GLTSPGSPGPDGKTGPPGAGQDGRP 300
 Db 758 GKDGVRLTGLPIGPPGAPAGDKGESGSPGAGTGAAGAPGDRGPPGPPGAPGAPGP 817
 QY 301 GPPGPPGARGO-----AGVMGPPGPKGAAGEPCKAGRGVPPGPA----- 341
 Db 818 GADGQPGAKGPPGADGAGKGDAGPFPAGPAGPFPPIGNVAPGAKGARGSAGSPGATGFP 877
 QY 342 -----VGPAGKDGAGAGAOQPPGAPG---AGERGEQPPAGSPGQGLPGGAPPPGAGKP 393
 Db 878 GAAGRVGPPGSGNAGPPGPPGAGKEGKGRGTGTGAPRGVEVPPGPPGAGEKSP 937
 QY 394 GEQGVPCDLGAPGSPGAGEPGTGLPQPPGGRGPGSRGPPGADGVAGPKPAGERGSP 453
 Db 938 GADGAPAGPFPGQGIAGQGVVGLPQGRGERGPPGLPGSPGPPGKQSGSGASGERGPP 997
 QY 454 GPAGP---KGSPPGAEGRPEAGLPCAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGP 510
 Db 998 GPMGPPGLAGPPESGRGAPGAEPSGDRGSGPANGRGETGAGPPGAPGAPGPV 1057
 QY 511 GARGQAGVGMGPPKPAAGEPCKAGRGVPPGPPGAVGPAKDGK-----AGAQ 558
 Db 1058 GPAGKSGDRGTGTPAGPAGPVGPAGARGPAGPQGRGDKGTGQDRIKGRHGSGLQ 1117
 QY 559 GPPGP-----AGPAGEREQCPAGSP---CFQGLPGAGPPGKAGKGEQGV 603
 Db 1118 GPPGPPGSGFSGSGSAGPAGPPGSGAGPAGKQGLNGLPGIPGPPGRGTGDGPV 1177
 QY 604 GDLGAPGSPGAP 616
 Db 1178 GPPGPPGPPGP 1190

1 GPP-----GEPTGLPGPPGSGRGGPGADGVAGPKPAGERSGPA 48
458 GPPGAGEGKRGARPGPTGLPPGSGRGGPGADGVAGPKPAGERSGPA 517
49 GPKSGEAGRPGEAGLPCAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 108
518 GPKSGEAGRPGEAGLPCAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 577
109 GVMGPPGPKGAAGEPGKAGRGVPPGAVGAGKAGAGAGAGAGAGAGAGAGAGAG 168
578 GVMGPPGPKGAAGEPGKAGRGVPPGAVGAGKAGAGAGAGAGAGAGAGAGAGAG 637
169 GSPGFGGLPGPAGPGEAGKPGEGVPGDLGAPGSGPAGE-----PGP---- 212
638 GSPGFGGLPGPAGPGEAGKPGEGVPGDLGAPGSGPAGE-----PGP---- 697
213 -----TCLPPGPERGSGRGGPGADGVAGPKPAGERSGPA 258
698 GANGAPNDGAKGADAGAPGAGSQGAPGLQGMFGERGAAGLPGPKGDRDAGPKGADGSP 757
259 GE-----AGRPGEAGLPCAK-----GLTSGPSGPDGKTGPPGAGQDGRP 300
758 GKDGVRLGTGPIGPPGAPADPKGESGSGPAGTGAAGACDRGEPGPPGAGFAGPP 817
301 GPPGPPGARGQ-----AGVMGPPGKGAAGEPGKAGRGVPPGPPGA----- 341
818 GADGPPGAKGEPDAGAKGADAGPAGPAGPPGPGIKNVAGPAGKAGSAGPPGATGPP 877
342 -----VGPAGKDEAGAGAGPPGAGP-----AGERGQAGSAGPPGPPGAGP 393
878 GAAGRVPGPPGSGNAGPPGPPGAGKGGKGRGTGTPAGRGVPPGPPGAGKGGSP 937
394 GEQGVPGDLGAPGPPGAGPAGTGLPPGPPGSGRGGPGADGVAGPKPAGERSGSP 453
938 GADGPPGAGTGTGPPGAGQGVVGLPGQGERGPPGLPGPSGEPGKQGPSAGSAGPP 997
454 GPAGP-----KSGPGEAGRPGEAGLPCAKGLTSGPSGPDGKTGPPGAGQDGRPPGPP 510
998 GVMGPPGAGLPGPGEAGKPGEGVPGDLGAPGSGPAGE-----PGP---- 1057
511 GARGQAGVMGPPGPKGAAGEPGKAGRGVPPGPPGAGVAGPAGKDE-----AGAQ 558
1058 GPAGSGDRGTGTPAGPAGPVGPPGAGPAGPQGRGDKGTGEGQDGRGKGRGFSGLQ 1117
559 GPPGP-----AGPAGERGQAGPAGP-----GFGGLPGPAGPGEAGKPGEQGV 603
1118 GPPGPPGSGEPGQGPSAGSAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1177
604 GDLGAPGSGPAG 616
1178 GPPGPPGPPGPPG 1190

RESULT 11
ADD45051
ID ADD45051 standard; protein; 1464 AA.
AC ADD45051;
XX
XX 29-JAN-2004 (first entry)
XX Human Protein P02452, SEQ ID NO 10483.
XX Human; pain; neuronal tissue; gene therapy;
XX spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.
OS Homo sapiens.
XX
XX WO2003016475-A2.
XX
XX 27-FEB-2003.
XX

RESULT 10
ADD45055
ID ADD45055 standard; protein; 1464 AA.
AC ADD45055;
XX
XX 29-JAN-2004 (first entry)
XX Human Protein P02452, SEQ ID NO 10487.
XX Human; pain; neuronal tissue; gene therapy;
XX spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.
XX
XX Homo sapiens.
XX
XX WO2003016475-A2.
XX
XX 27-FEB-2003.
XX
XX 14-AUG-2002; 2002WO-US025765.
XX
XX 14-AUG-2001; 2001US-0312147P.
XX 01-NOV-2001; 2001US-0346382P.
XX 26-NOV-2001; 2001US-0333347P.
XX
XX (GEMO) GEN HOSPITAL CORP.
XX (FARB) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
XX
XX WPI; 2003-268312/26.
XX GENBANK; P02452.
XX
XX New composition comprising two or more isolated polypeptides, useful for
XX preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
XX or human polynucleotides or a polynucleotide which represents a fragment,
XX derivative or allelic variation of the nucleic acid sequence. Also
XX claimed are a vector comprising the novel polynucleotide, a host cell
XX comprising the vector, a method for identifying a nucleotide sequence
XX which is differentially regulated in an animal subjected to pain and a
XX kit to perform the method, an array, a method for identifying an agent
XX that increases or decreases the expression of the polynucleotide sequence
XX that is differentially expressed in neuronal tissue of a first animal
XX subjected to pain, a method for identifying a compound which regulates
XX the expression of a polynucleotide sequence which is differentially
XX expressed in an animal subjected to pain, a method for identifying a
XX compound that regulates the activity of one or more of the
XX polynucleotides, a method for producing a pharmaceutical composition, a
XX method for identifying a compound or small molecule that regulates the
XX activity in an animal of one or more of the polypeptides given in the
XX specification, a method for identifying a compound useful in treating
XX pain and a pharmaceutical composition comprising the one or more
XX polypeptides or their antibodies. The polynucleotide or the compound that
XX modulates its activity is useful for preparing a medicament for treating
XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
XX therapy). The sequence presented is a human protein (shown in Table 2 of
XX the specification) which is differentially expressed during pain. Note:
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic form directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.

Seq Sequence 1464 AA;
Query Match 64.1%; Score 2237.5; DB 7; Length 1464;
Best Local Similarity 60.2%; Pred. No. 3.9e-130;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

PT vaccine, and for treating autoimmune disorders, infections and cancer.
PS Example 3; Fig 8; 168pp; English.
XX
CC The present sequence is porcine alpha(I) collagen. The present invention
CC relates to recombinant synthesis of collagens and gelatins derived from
CC animals. Collagen is useful in medical, pharmaceutical, food and cosmetic
CC industries. Collagen is an important component of arterial sealants, bone
CC grafts, drug delivery system, dermal implants, haemostats, and
CC incontinence implants, and for treating autoimmune disorders such as
CC rheumatoid arthritis. Collagen is useful in food products such as sausage
CC casings, and in cosmetics or facial and skin products such as
CC moisturisers. Recombinant gelatin is useful in vaccine formulations for
CC treating viral infections, autoimmune diseases and cancer. Gelatin is
CC useful in the manufacture or as a component of various pharmaceutical and
CC medical devices and products, in food and beverage industries, in hair
CC care and skin care products, as a glue or adhesive in various
CC manufacturing processes, as a light-sensitive coating in various
CC electronic devices, as photoresist base in photolithographic processes,
CC in printing and photographic applications, in laboratory application, and
CC as a component in various gels used for biochemical and electrophoretic
CC analysis, including enzymographic gels
XX
SQ Sequence 1449 AA;

Query Match 64.0%; Score 2234; DB 4; Length 1449;
Best Local Similarity 62.0%; Pred. No. 6.4e-130;
Matches 438; Conservative 21; Mismatches 154; Indels 94; Gaps 13;
QY 1 GPPGPGTGLGPPGGRGGSGRFGADGAVGPKGAGRGSPGAGKSGPGEAGRP 60
Db 472 GARGGGAGLPGPPGGRGGSGRFGADGAVGPKGAGRGSPGAGKSGPGEAGRP 531
QY 61 GEAGLPGAKGLTSGSPGPPDGKTPGPPAGQDGRPPGPPGARGQAGVGMGPPGKGA 120
Db 532 GEAGLPGAKGLTSGSPGPPDGKTPGPPAGQDGRPPGPPGARGQAGVGMGPPGKGA 591
QY 121 GEPKAGRGVPPGAVGPKAGKDEAGACQPPGAPGAGRGQAGSPGGLPGPA 180
Db 592 GEPKAGRGVPPGAVGPKAGKDEAGACQPPGAPGAGRGQAGSPGGLPGPA 651
QY 181 GPPGAGKFGQGVGPDGLGAPGSPGAGE-----PGP----- 212
Db 652 GPPGAGKFGQGVGPDGLGAPGSPGAGE-----PGP----- 711
QY 213 --TGLPDPGRRGGSGRFGADGAVGPKGAGRGSPGAGKSGPGE-----A 261
Db 712 GDAGAPAGSQGAPGLQGMPPGRRGAAGLPGPKDGRDAGPKADGAPGKGVRLTGP 771
QY 262 GRPGAGLPGAKGLTSGSPGPPDGKTPGPPAGQDGRPPGPPGARGQAGVGMG 317
Db 772 GPPGAGLPGAKGLTSGSPGPPDGKTPGPPAGQDGRPPGPPGARGQAGVGMG 831
QY 318 ---PQPKGAAGEPKAGRGVPPGPA-----VGPAGKGEAGACQPPGAPGAGE 365
Db 832 TGPPGPIGVCAPGPKGARGAGPPGATGFPAGAGRVGPPGSPGNAGPPGPPGAGKEG 891
QY 366 ---RGEQGPAGSPFQGLPGPAGPBGAGKGEQGVGDLGAPDPSGAPGEPGPTGLPG 422
Db 892 KGRPGETPAGRPGGAGPPGPPGAGKSGSPGADGAPGAFPTPGPQIAGQGVVGLPG 951
QY 423 PGEKGGPSRGFPAGDGVAGPKGAPAGRGSPGAPKSGSPGACRPGGAGLPGAKGLTGS 482
Db 952 RGERGFPGLPFSGEPGKQSGSPGREGPPGMPG---PGLAGPPGSRGAPGAGS 1008
QY 483 P---GSPGPD---GKTGPPGAGQDGRPPGPPGAGQAGVGMGFPGPKGAAGBPGKAGE 536
Db 1009 FGRDCAFPKGRGSGPAGPPGAPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAP 1068
QY 537 RGVPGPAGVAPGAKGGE-----AGAGQPPG-----ACAPGERGE 572
Db 1069 RGPAGPQGRDQKGETGEQGRGKTHGRGSLGQPPGPPGSPGQEGSPGASGAPGRGP 1128

QY 573 QGPAGSP---GFGGLPGPAGPPEAGKPGCEQGVGPDGLGAPGSPGAG 616
Db 1129 PGAGAPGKXGLNGLPGIPGPPGRTGCDAGPVGPPGPPGPPGPPG 1175
RESULT 13
ID ADE87050
XX ADE87050 standard; protein; 1161 AA.
AC ADE87050;
XX
DT 29-JAN-2004 (first entry)
XX
XX Human pancreatic cell protein sequence SeqID510.
XX neoplastic pancreatic cell; pancreatic cancer;
XX cancer death; cytostatic; vaccine; Gene therapy;
XX non-cancerous pancreas disease; human.
OS Homo sapiens.
XX
XX WO2003060145-A2.
XX
XX 24-JUL-2003.
XX
XX 19-DEC-2002; 2002WO-US040655.
XX
XX 21-DEC-2001; 2001US-0342768P.
XX
XX (DIAD-) DIADEXUS INC.
XX
XX Sun Y, Liu C;
XX
XX WPI; 2003-587286/55.
XX
XX N-PSDB; AD387387.
XX
XX New pancreatic specific nucleic acid molecule or protein for diagnosing,
XX staging, imaging, monitoring, preventing or treating pancreatic cancer or
XX non-cancerous disease states of the pancreas.
XX
XX Claim 12; SEQ ID NO 510; 635pp; English.
XX
XX This invention relates to novel nucleic acids and proteins present in
XX normal and neoplastic pancreatic cells. Pancreatic cancer is a common
XX cause of cancer death worldwide, therefore accurate methods of diagnosis
XX and treatment are required. Compounds which modulate the proteins of the
XX invention may have cytostatic activity and the protein and DNA sequences
XX of the invention may be useful for the development of a vaccine or in
XX gene therapy. The composition and methods are useful in diagnosing,
XX staging, imaging, monitoring, preventing or treating pancreatic cancer
XX and non-cancerous disease states of the pancreas. The present sequence is
XX that of a human pancreatic protein of the invention.
XX
SQ Sequence 1161 AA;
Query Match 64.0%; Score 2233.5; DB 7; Length 1161;
Best Local Similarity 60.0%; Pred. No. 5.7e-130;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;
QY 1 GPP-----GEPGPTGLPQPPGGRGGSGRFGADGAVGPKGAGRGSPGPA 48
Db 155 GPPGAGBEGKRGARPEGPTGLPQPPGGRGGSGRFGADGAVGPKGAGRGSPGPA 214
QY 49 GPKGSPGAGRGAGLPGAKGLTSGSPGPDGKTGPPGAGODGRGPPGPPGARGQA 108
Db 215 GPKGSPGAGRGAGLPGAKGLTSGSPGPDGKTGPPGAGODGRGPPGPPGARGQA 274
QY 109 GYWGPPGPKGAAGEPKAGRGVPPGAVGPKDGEAGACQPPGAPGAGPAGERGEQGPA 168
Db 275 GYWGPPGPKGAAGEPKAGRGVPPGAVGPKDGEAGACQPPGAPGAGPAGERGEQGPA 334
QY 169 GSPGQGLPDPAGPGEAGKPGCEQGVGPDGLGAPGSPGAGE-----PGP----- 212

Db 335 GSPGQGLPGPAGPPGAGKPEQGVPGDLGAPGSPGARGGPPGGERGVQGGPPGAPGR 394
 QY 213 -----TGLPGPPGERGGSGRGGFPAGDVAGPKGPPAGRGSGPPGAGKGGSP 258
 Db 395 GAGAPGNDGAKGADAGAPGAFGSGQGLQMPGREGAAGLPKPGKDRGDAGPKGADGSP 454
 QY 259 GE-----AGRPGEAGLPKAK-----GLTSGPSGPPGDKTGPAGQDGRP 300
 Db 455 GKDGVRGLTGPIGPPGAGAGDKGSGSPGAPGTGARGAPCDRGEPGPPGAGPAP 514
 QY 301 GPPGPPGARGO-----AGVMGPPGKGAAGEPKGAGRGVGPVPPGA----- 341
 Db 515 GADGQPKAGGPPGADGAKGADGPPGAPGAPGPPGFIQNVGAPKAGARGAGPPGATGFP 574
 QY 342 -----VGPAGKDGAGAGPPGAPG---AGERGEQGPAGSPGQGLPGPAGPPGAGKP 393
 Db 575 GAAGRVGPPGSGNAGPPGPPGAPGKGGKGGPRGTGPRGPEVGPVPPGPPGAGEKGP 634
 QY 394 GEQGVPGDLGAPGSPGAPGPPGTLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 453
 Db 635 GADGAPAGPPTGPPGQTAGQGVVGLPQGRGERGFPGLPGSGEPGKQGPSGASGERGPP 694
 QY 454 GPAGP---KSPGAGRPPGAGLPGAKLTGSPGSPGPPGPPGPPGPPGPPGPPGPPGPP 510
 Db 695 GPMGPPGLAGPPGSGREGAPGAGSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 754
 QY 511 GARGQGVWGPFGPKGAAGEPKGAGRGVPPGPPGAVGPPGAKDGE-----AGAQ 558
 Db 755 GPAGKSGDRGTGTPAGPAGPVGVPVAGPAGPQPGKGTGEGQDGRGKGRHGFSLQ 814
 QY 559 GPPGP-----AGPAGERGEQGPAGSP---CFQGLPGPAGPPGAGPYGQGV 603
 Db 815 GPPGPPGSGPQSPGSGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 874
 QY 604 GDLGAPGSPGAPG 616
 Db 875 GPPGPPGPPGPPG 887

RESULT 14

ABG93947
 ID ABG93947 standard; protein; 1461 AA.

XX AC ABG93947;

XX DT 26-NOV-2002 (first entry)

XX DE Human polypeptide orthologous to DACC-11.

XX KW Human; deer; rat; mouse; DACC; deer antler cartilage cell;
 KW cell stimulation; cell inhibition; cell growth; cell division;
 KW mesenchymal cell; chondrocyte; chondrogenesis; osteogenesis; growth;
 KW repair; regeneration; restoration; extracellular matrix;
 KW cartilaginous matrix; cartilage; disc; connective tissue; agonist;
 KW antagonist; gene therapy.

XX OS Homo sapiens.

XX FN W0200264625-A1.

XX PD 22-AUG-2002.

XX PF 15-FEB-2002; 2002WO-AU000163.

XX PR 15-FEB-2001; 2001AU-00003116.

XX XX (ADPP-) ADP PHARM PTY LTD.

XX PA (UNSY) UNIV SYDNEY.

XX XX Roubin R, Ghosh P;

XX PI WPI; 2002-643456/69.

XX DR WPI; 2002-643456/69.

PT Stimulating or inhibiting cell growth and/or division, useful for
 PT stimulating chondrogenesis, cartilage, disc or connective tissue growth,
 PT repair, and/or regeneration, comprises administering deer antler
 PT cartilage gene.

XX Claim 13; Page 154-160; 214pp; English.

XX The invention discloses a method for stimulating or inhibiting cell
 CC growth and/or division which comprises contacting or inserting into an
 CC animal cell a polypeptide comprising one of the deer antler cartilage
 CC cell (DACC) clones disclosed. More particularly, the method relates to
 CC these polypeptides stimulating mesenchymal cell growth and/or division
 CC and to transfecting these cells and chondrocytes with vectors carrying
 CC the genes of these polypeptides capable of stimulating chondrogenesis,
 CC osteogenesis, growth, repair, regeneration and/or restoration of the
 CC extracellular matrix. The chondrocytes selectively express genes required
 CC to form a cartilaginous matrix. The DACC polypeptides and polynucleotides
 CC are useful for identifying an agent that modulates the activity of the
 CC polypeptide, for stimulating mesenchymal cell growth and/or division by
 CC exposing animal mesenchymal cells to conditioned media or its active
 CC fraction, obtained from deer antler cartilage cells, for inhibiting cell
 CC growth and/or division by inserting into an animal cell, a compound which
 CC inhibits the translation of the polynucleotide encoding the DACC. The
 CC method and the polypeptides are useful for stimulating mesenchymal cell
 CC growth and/or division or for stimulating chondrogenesis, cartilage, disc
 CC or connective tissue growth, repair, regeneration and/or restoration in
 CC an animal. The polynucleotides, polypeptides, agonists and antagonists
 CC may be used in treatment modalities, specifically in gene therapy. The
 CC polypeptides can be used as bait proteins in a two- or three-hybrid assay
 CC to identify other proteins, which bind to or interact with the
 CC polypeptide and are involved in modulating cell growth and/or division.
 CC The sequences presented in ABG93923-ABG93948 are the proteins encoded by
 CC the DACC cDNA clones

XX Sequence 1461 AA;

Query Match 64.0%; Score 2233.5; DB 5; Length 1461;
 Best Local Similarity 60.0%; Pred No. 6.9e-130;
 Matches 440; Conservative 23; Mismatches 133; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPFPGERGSGRFPFGADGVAGPKGAGERGSGP 48
 Db 455 GPPGPAEGKRGARGEPGTLPGPPGERGSGRFPFGADGVAGPKGAGERGSGP 514
 QY 49 GPKSGPGEAGRPGEAGLPKAKLTGSPGSPGPPGPPGPPGPPGPPGPPGPPGPP 108
 Db 515 GPKSGPGEAGRPGEAGLPKAKLTGSPGSPGPPGPPGPPGPPGPPGPPGPPGPP 574
 QY 109 GVMGPPGPKGAAGEPKGAGRGVPPGPAVGPAGKDGEGAGAQGPPGAPAGERGEGCPA 168
 Db 575 GVMGPPGPKGAAGEPKGAGRGVPPGPAVGPAGKDGEGAGAQGPPGAPAGERGEGCPA 634
 QY 169 GSPGQGLPGPAGPPGAGKGEQGVPGDGLGAPGSPGAPG-----RGP----- 212
 Db 635 GSPGQGLPGPAGPPGAGKGEQGVPGDGLGAPGSPGAGRGVPPGPPGAPGR 694
 QY 213 -----TGLPGPPEGERGSGRFPFGADGVAGPKGAGERGSGPAGPKGSP 258
 Db 695 GAGAPGNDGAKGADAGAPGSGQAGPQLQMPGREGAAGLPKPGKDRGDAGPKGADGSP 754
 QY 259 GE-----AGRPGEAGLPKAK-----GLTSGPSGPPGDKTGPAGQDGRP 300
 Db 755 GKDGVRGLTGPIGPPGAGAGDKGSGSPGAPGTGARGAPCDRGEPGPPGAGPAP 814
 QY 301 GPPGPPGARGO-----AGVMGPPGKGAAGEPKGAGRGVGPVPPGA----- 341
 Db 815 GADGQPKAGGPPGADGAKGADGPPGAPGAPGPPGFIQNVGAPKAGARGAGPPGATGFP 874
 QY 342 -----VGPAGKDGAGAGPPGAPG---AGERGEQGPAGSPGQGLPGPAGPPGAGKP 393
 Db 875 GAAGRVGPPGSGNAGPPGPPGAPGKGGKGGPRGTGPRGPEVGPVPPGPPGAGEKGP 934
 QY 394 GEQGVPGDLGAPGSPGAPG 616
 Db 875 GPPGPPGPPGPPG 887

935 GADGAGAGTGGQGIAGQGVVGLFGQGERGFPGLPGSPGPGKQSGASGERGPP 994
454 GPAGP---KGSFGAGPGEAGLPGAKGLTGSPGSPGDKTGPAGQDGRPPGPP 510
995 GPMGPPGLAGPSPGSGREGAPGARGSPGCDGSPGAKGDRGTGPAGPAGPAGPVP 1054
511 GARGOAGVMGPPGPKGAAGEPKAGRGVPGPPGAVGPAKDGSP-----AGAQ 558
1055 GPAGKSGDRGTGPAGPAGPVPVVGARGPAGPQPRGDKGTSGQDRG:KGHRGFSGLQ 1114
559 GPPGP-----AGPAGERGEOGPAGSP---GFOGLPGPAGPAGPAGKPGEGVVP 603
1115 GPPGPPGSPGSPGSGASGAPRPGSPGAGAPKDGGLGILPGPIGPPGPRGTGDGPV 1174
604 GDLGAPGSPGAG 616
1175 GPPGPPGPPGPPG 1187

RESULT 15
AAU14136
ID AAU14136 standard; protein, 1464 AA.
AC AAU14136;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human novel protein #7.
XX
KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnary; nocotropic;
KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW parkinson's disease; immunogen; antibody; gene therapy; neurological disorder;
KW thrombolytic; immunogen; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX
OS Homo sapiens.
XX
FN WO200155437-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US002623.
XX
PR 25-JAN-2000; 2000US-00491404.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-451939/48.
XX
DR N-PSDB; AAS22441.
XX
Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage.
XX
PS Example 4; Page 525-527; 894pp; English.

The invention relates to polynucleotides encoding novel human proteins or
their active domains. The polypeptides, polynucleotides and antibodies
raised against the polypeptides are used in a method of treatment of a
mammal and prevention of disorders caused by the aberrant protein
expression or activity. The polypeptides can be used as molecular weight
markers, food supplements, and in antibody production. The polypeptides
are used to identify compounds which bind to the polypeptides.
Polynucleotides of the invention are used as probes and primers, for
sequencing, for chromosome or gene mapping, in the production of
recombinant proteins, and in generating anti-sense DNA or RNA and in gene
therapy. Polypeptides of the invention can be used to target drugs to a
tumour, in assays to determine biological activity, to raise
antibodies/elicit an immune response, to determine quantitative protein

CC levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
disorders, stem cell disorders, regenerating bone, cartilage, tendon,
ligament and/or nerve tissue, wound healing, treating burns, promoting
the proliferation, differentiation and survival of stem cells, as a
contraceptive, treating osteoporosis and osteoarthritis, anaemia,
Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
diseases, nervous system disorders, and infection. The present sequence
CC represents a protein of the invention
XX
SQ Sequence 1464 AA;

Query Match 64.0%; Score 2233.5; DB 4; Length 1464;
Best Local Similarity 60.0%; Pred. No. 6.9e-130;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;
QY 1 GPP-----GEPGPTGLPQPPGERSGSPGFPAGDVAGPKPAGERGSPGA 48
DB 458 GPPGAGEBKGARGEPGPTGLPQPPGERSGSPGFPAGDVAGPKPAGERGSPGA 517
QY 49 GPKSGPENGRRGEAGLPGAKGLTCSGSPGPDGKTGPRGAGODRGPGRPGARCOA 108
DB 518 GPKSGPENGRRGEAGLPGAKGLTCSGSPGPDGKTGPRGAGODRGPGRPGARCOA 577
QY 109 GVMGPPGPKGAAGEPKAGRGVPGPPGAVGPAKDGAGAGQPPGPPGAPAGERGEOGPA 168
DB 578 GVMGPPGPKGAAGEPKAGRGVPGPPGAVGPAKDGAGAGQPPGPPGAPAGERGEOGPA 637
QY 169 GSGFGQLGPPGAPGEGEAGKPOEQVPGDLGAPGSSPAGE-----PGP---- 212
DB 638 GSGFGQLGPPGAPGEGEAGKPOEQVPGDLGAPGSSPAGE-----PGP---- 697
QY 213 -----TGLPQPPGERSGSPGFPAGDVAGPKPAGERGSPGAPKPGSP 258
DB 698 GARGAPNDGAKGDAGAPGAPGSPQAGLQGVGGERGAAGLPGKDRGDAGPKGADGSP 757
QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGSPGPKTGTGPPGAGDGRP 300
DB 758 GKDGVRLGTGTPGPPGAPAGPDKGSGSPGAPGTGARGAPGDRGEPGPPGAPGAP 817
QY 301 GPPGPPGARGQ-----AGVMGPPGPKGAAGEPKAGRGVPGPPGA---- 341
DB 818 GADGQPKAGGEPGDGAKGDAGPDPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPP 877
QY 342 -----VGPAGKDGEGAGAGQPPGAPG---AGERGEGPAGSPGFGQLGPPGAPG 393
DB 878 GAAGRVGPPGSPGAGPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 937
QY 394 GSGQVPGDLGAPGSPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 453
DB 938 GADGAPAGPFTGPGQGIAGQGVVGLFGQGERGFPGLPGSPGPGKQSGASGERGPP 997
QY 454 GPAGP---KGSFGAGPGEAGLPGAKGLTGSPGSPGPDGKTGPPGAGDGRPPGPP 510
DB 998 GPMGPPGLAGPSPGSGREGAPGARGSPGCDGSPGAKGDRGTGPAGPAGPAGPVP 1057
QY 511 GARGOAGVMGPPGPKGAAGEPKAGRGVPGPPGAVGPAKDGSP-----AGAQ 558
DB 1058 GPAGKSGDRGTGPAGPAGPVPVVGARGPAGPQPRGDKGTSGQDRG:KGHRGFSGLQ 1117
QY 559 GPPGP-----AGPAGERGEOGPAGSP---GFOGLPGPAGPAGPAGKPGEGVVP 603
DB 1118 GPPGPPGSPGSPGSGASGAPRPGSPGAGAPKDGGLGILPGPIGPPGPRGTGDGPV 1177
QY 604 GDLGAPGSPGAG 616
DB 1178 GPPGPPGPPGPPG 1190

Search completed: October 18, 2004, 13:37:26

Tue Oct 19 09:27:57 2004

us-10-658-989a-2.rag

Page 14

Job time : 104.398 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:31:20 ; Search time 24.4569 Seconds
(without alignments)
1673.075 Million cell updates/sec

Title: US-10-658-989A-2

Perfect score: 3488
Sequence: 1 GPPGEPGPTCLPFPGERGG.....GEGVFDLGAQPSGAGG 617

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2237.5	64.1	1464	4	US-09-331-347C-21
2	2228.5	63.9	1341	3	US-08-963-825-18
3	2228.5	63.9	1341	3	US-09-500-811-18
4	2228.5	63.9	1341	3	US-09-570-573-18
5	2228.5	63.9	1341	3	US-09-548-608-18
6	2224.5	63.8	1461	4	US-09-585-887-9
7	2224.5	63.8	1461	4	US-09-289-578-9
8	2208.5	63.3	1057	3	US-08-931-820-1
9	2191	62.8	822	3	US-09-219-849-49
10	1955	56.0	1017	4	US-08-468-986-10
11	1943	55.7	1060	3	US-08-931-820-3
12	1943	55.7	1418	3	US-08-963-825-20
13	1943	55.7	1418	3	US-09-010-999-1
14	1943	55.7	1418	3	US-09-500-811-20
15	1943	55.7	1418	3	US-09-570-573-20
16	1943	55.7	1418	3	US-09-548-608-20
17	1936	55.5	1442	2	US-08-316-650-12
18	1936	55.5	1442	5	PCT-US95-02251-12
19	1882.5	54.0	595	3	US-09-219-849-48
20	1882.5	54.0	595	3	US-09-219-849-50
21	1882.5	54.0	1064	1	US-08-642-255-62
22	1853	53.1	1057	3	US-08-931-820-4
23	1837	52.7	720	3	US-09-219-849-4
24	1837	52.7	777	1	US-08-642-255-53
25	1833.5	52.6	1078	3	US-08-963-825-21
26	1833.5	52.6	1078	3	US-09-500-811-21
27	1833.5	52.6	1078	3	US-09-570-573-21

28 1833.5 52.6 1078 3 US-09-548-608-21 Sequence 21, Appli
29 1764 50.6 1024 3 US-08-931-820-2 Sequence 2, Appli
30 1764 50.6 1366 3 US-08-963-825-19 Sequence 19, Appli
31 1764 50.6 1366 3 US-09-500-811-19 Sequence 19, Appli
32 1764 50.6 1366 3 US-09-570-573-19 Sequence 19, Appli
33 1764 50.6 1366 3 US-09-548-608-19 Sequence 10, Appli
34 1764 50.6 1366 4 US-09-585-887-10 Sequence 10, Appli
35 1762 50.5 1366 4 US-09-289-578-10 Sequence 12, Appli
36 1699.5 48.7 1065 1 US-08-642-255-72 Sequence 5, Appli
37 1674 48.0 960 3 US-09-219-849-5 Sequence 12, Appli
38 1645 47.2 492 4 US-08-468-986-12 Sequence 56, Appli
39 1644 47.1 1806 4 US-09-919-497-56 Sequence 132, App
40 1619.5 46.4 829 1 US-08-642-255-132 Sequence 53, Appli
41 1619.5 46.4 829 1 US-08-397-633A-53 Sequence 68, Appli
42 1619.5 46.4 837 1 US-08-175-155-68 Sequence 103, App
43 1619.5 46.4 837 1 US-08-477-509B-103 Sequence 101, App
44 1619.5 46.4 837 1 US-08-642-255-101 Sequence 75, Appli
45 1619.5 46.4 837 2 US-08-707-237A-75

ALIGNMENTS

RESULT 1

US-09-331-347C-21
; Sequence 21, Application US/09331347C
; Patent No. 6617431
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, a
; FILE REFERENCE: 1149-3
; CURRENT APPLICATION NUMBER: US/09/331,347C
; CURRENT FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-331-347C-21

Query Match 64.1%; Score 2237.5; DB 4; Length 1464;
Best Local Similarity 60.2%; Pred. No. 1.2e-133;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPQPPGGRGSGRGGADGAGVAGPKGPGAGERSGPGA 48
Db 458 GPPGAGEGKRGARCEPCTCLPQPPGGRGSGRGGADGAGVAGPKGPGAGERSGPGA 517
QY 49 GKSGPGEAGRGEGAGLPGAKLTGSPGSPGDKTGTGPPGAGODGRGPPGARGQA 108
Db 518 GKSGPGEAGRGEGAGLPGAKLTGSPGSPGDKTGTGPPGAGODGRGPPGARGQA 577
QY 109 GWMGPPGPKGAAGEGKAGRGVPPGAVGAGKDGAGAGAGGPPGAGPAGERGEGQSPA 168
Db 578 GWMGPPGPKGAAGEGKAGRGVPPGAVGAGKDGAGAGAGGPPGAGPAGERGEGQSPA 637
QY 169 GSPGFQGLPAGPAGPGEAGKPGEQVPGDLGAPGSGPAGE-----PGP----- 212
Db 638 GSPGFQGLPAGPAGPGEAGKPGEQVPGDLGAPGSGPAGE-----PGP----- 697
QY 213 -----TCLPQPPGGRGSGRGGADGAGVAGPKGPGAGERSGPGAPKGGSP 258
Db 698 GANGAPNDGAKGDAGAPGAPGSGAGLPGKDRDAGKGDGSP 757
QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGPPGDKTGTGPPGAGODGRP 300
Db 758 GKDVRLGTCPTGPPGAGAGPDKGSGSPGAGPTGARGAPCDRGEPPGPPGAGPAGPP 817
QY 301 GPPGPPGARGQ-----AGVMGPPGKGAAGEGKAGRGVPPGPGA---- 341


```

; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
; US-09-500-811-18

Query Match 63.9%; Score 2228.5; DB 3; Length 1341;
Best Local Similarity 60.0%; Pred. No. 4.2e-133;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GERPGTGLPGPGRGPGSRGPPGADGVAGPKGAGRGSGPPA 48
DB 334 GPPGAGEGKRGARGEPGPTGLPGPGRGPGSRGPPGADGVAGPKGAGRGSGPPA 393
QY 49 GPKGSPGAGPGGAGLPGCAKGLTSGSPGPDGKTGTPGAGQDGRPPGPPGARGQA 108
DB 394 GPKGSPGAGPGGAGLPGCAKGLTSGSPGPDGKTGTPGAGQDGRPPGPPGARGQA 453
QY 109 GVMGPPGPKGAAGFPGKAGRGVPPGAVGPAKGDSAGAGQPPGAGPAGERGQGPA 168
DB 454 GVMGPPGPKGAAGFPGKAGRGVPPGAVGPAKGDSAGAGQPPGAGPAGERGQGPA 513
QY 159 GSPGQGLPGAGPPGEGAGKEGQVPGDLGAPGSPGAGE-----PGP--- 212
DB 514 GSPGQGLPGAGPPGEGAGKEGQVPGDLGAPGSPGAGE-----PGP--- 573
QY 213 -----TGLPGPGRGPGSRGPPGADGVAGPKGAGRGSGPPGAGXGSP 258
DB 574 GAGAPGNDGAKGDAGAFGAGSQAGLQGVPGERGAAGLPGPKGDRGDAGPKGADGSP 633
QY 259 GE-----AGRGEAGLPGAK-----GLTSPGSPGPDGKTGTPGAGQDGRP 300
DB 634 GKDGVRLGTGTPGPPGAPGDKGESGPPGAGTGAAGAPGDRGEPGPPGAGFAGPP 693
QY 301 GPPGPPGARGO-----AGVMGPPGPKGAAGFPGKAGRGVPPGPA----- 341
DB 694 GADGQFGAKGEPGDAGAKDAGPPGAPGAPGPPGIGNVAPGAKGARGSPGATGFP 753
QY 342 -----VGPAGKDGAGAGQPPGAGP-----AGERGQGPAGSFGGGLPGPAGPPGEGAKP 393

```

RESULT 4

```

US-09-570-573-18
; Sequence 18, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Ovist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)

```

```

DB 754 GAAGRVGPPGSGNAGPPGPGAGKEGKGRGTGTPAGRFGVGVPPGPPGAGEKGGSP 813
QY 394 GEQGVPGDLGAPGSPGAGGEGPTGLPGPGRGPGSRGPPGADGVAGPKGAGRGSGP 453
DB 814 GADGAPAGPPTPGQGIAGQGVVGLPQGRGERFPGLPGSPGPKGKSGASGERGPP 873
QY 454 GPAGP---KSPGGEAGRFEGAGLPGCAKGLTSGSPGPDGKTGTPGAGQDGRPPGPP 510
DB 874 GPMGPPGLAGPPGSGREGAPGAEKSPGDRGSPGAKGDRGTGTPAGPFGAGXGAGP 933
QY 511 GARGQAGVMGPPGPKGAGRGVPPGAGVGPAGPKDGE-----AGAQ 558
DB 934 GPAGKSGDRGTGTPAGPAGVGPAGPAGPQGRGDKGTGSGDGRGIRGKHGFGSLQ 993
QY 559 GPPGP-----AGPAGERGQGPAGSP---GFGGLPGPAGPFGGAGKPGEQGVP 603
DB 994 GPPGPPGSPGSGGASGSPGPPGAGPKGDLGNGLPGPPIGPPGPRGRTGDAGPV 1053
QY 604 GDLGAPGSPGAG 616
DB 1054 GPPGPPGPPGPPG 1066

```

US-09-570-573-18

Query Match 63.9%; Score 2228.5; DB 3; Length 1341;
 Best Local Similarity 60.0%; Pred. No. 4.2e-133;
 Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

```

QY 1 GPP-----GEPGPTGLPCCRGCGSRGPPGADGVAGPKGAGERGSGCPA 48
  |||
Db 334 GPPGAGBEGKRGARGEGPTGLPCCRGCGSRGPPGADGVAGPKGAGERGSGCPA 393
  |||
QY 49 GPKGSGBAGRPGEAGLPGAKGLTGSPPGDPDKTGTGPPAGDQGRPPGPPGARGQA 108
  |||
Db 394 GPKGSGBAGRPGEAGLPGAKGLTGSPPGDPDKTGTGPPAGDQGRPPGPPGARGQA 453
  |||
QY 109 GVMGFPKGAAGBPKGAGERGVPPGAVGPKAGDKGAGAGQPPGPPGAGERGEGQA 168
  |||
Db 454 GVMGFPKGAAGBPKGAGERGVPPGAVGPKAGDKGAGAGQPPGPPGAGERGEGQA 513
  |||
QY 169 GSPFGQLGPPGPPGAGPKGEGQVPGDLGAPGSPGAGE-----PGP--- 212
  |||
Db 514 GSPFGQLGPPGPPGAGPKGEGQVPGDLGAPGSPGAGERGPPGPPGPPGPPGPPG 573
  |||
QY 213 -----TGLPCCRGCGSRGPPGADGVAGPKGAGERGSGCPA 616
  |||
Db 1054 GPPGPPGPPGPPG 1066
  |||

```

RESULT 5

US-09-548-608-18
 ; Sequence 18, Application US/09548608
 ; Patent No. 6355442
 ; GENERAL INFORMATION:
 ; APPLICANT: Ovist, Per
 ; APPLICANT: Bonde, Martin
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 ; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC

STREET: 805 Third Avenue
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10022
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09548,608
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/187,319
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Gogolis, Adda C
 REGISTRATION NUMBER: 29,714
 REFERENCE/DOCKET NUMBER: 4305/08701
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7700
 TELEFAX: 212-753-6237
 TELEX: 236687
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1341 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: COLLAGEN ALPHA 1 (I)
 US-09-548-608-18

Query Match 63.9%; Score 2228.5; DB 3; Length 1341;
 Best Local Similarity 60.0%; Pred. No. 4.2e-133;
 Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

```

QY 1 GPP-----GEPGPTGLPCCRGCGSRGPPGADGVAGPKGAGERGSGCPA 48
  |||
Db 334 GPPGAGBEGKRGARGEGPTGLPCCRGCGSRGPPGADGVAGPKGAGERGSGCPA 393
  |||
QY 49 GPKGSGBAGRPGEAGLPGAKGLTGSPPGDPDKTGTGPPAGDQGRPPGPPGARGQA 108
  |||
Db 394 GPKGSGBAGRPGEAGLPGAKGLTGSPPGDPDKTGTGPPAGDQGRPPGPPGARGQA 453
  |||
QY 109 GVMGFPKGAAGBPKGAGERGVPPGAVGPKAGDKGAGAGQPPGPPGAGERGEGQA 168
  |||
Db 454 GVMGFPKGAAGBPKGAGERGVPPGAVGPKAGDKGAGAGQPPGPPGAGERGEGQA 513
  |||
QY 169 GSPFGQLGPPGPPGAGPKGEGQVPGDLGAPGSPGAGE-----PGP--- 212
  |||
Db 514 GSPFGQLGPPGPPGAGPKGEGQVPGDLGAPGSPGAGERGPPGPPGPPGPPGPPG 573
  |||
QY 213 -----TGLPCCRGCGSRGPPGADGVAGPKGAGERGSGCPA 616
  |||
Db 574 GAGAPGNDGAKGADAGAPGAPGSGQAPGLQMPGGERGAAGLPGPKGDRGADGPKGADGSP 633
  |||
QY 259 GE-----AGRPGEAGLPGAK-----GLTGSPPGSPGDPDKTGTGPPGPPGPPG 300
  |||
Db 634 GKDGVRLTGTGIPGPPGAGAPGDKGSGSPGAPGTGARGAPGDRGPPGPPGPPGPPG 693
  |||
QY 301 GPPGPPGARGO-----AGVMGFPKGAAGBPKGAGERGVPPGPPG 341
  |||
Db 694 GADQPGAKGPPGADGAGDAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 753
  |||
QY 342 -----VGPAGKDGAGAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 393
  |||
Db 754 GAAGRVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 813
  |||

```


QY 394 GEOVPGDLAGPSPGAGPFGPTGLPGPPGRRGSGRGGPAGDVGAPKGPAGRRGSP 453
Db 814 GADGAPAGTPTGPGQIAGQGVVGLPGQSGRGGPGLPGPSGEPKQSGASGERGPP 873
QY 454 GPAGP---KSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGDGRPPGPP 510
Db 874 GPMGPPGLAGPPGREGAGPAGAGSGSPGRCGSGAKGDRGETGAPGPGXGXGAPGPV 933
QY 511 GARGQAGVMGPPGPKGAAGPFGKAGRGVPGPPGAVGPAKDGEB-----AGAQ 558
Db 934 GPAGKSDRGRTGTPAGPAGVPGAGARGPAGPQCPRGDKGTGEGQDGRGKIHGRGFSGLQ 993
QY 559 GPPGPPGSGPQSGASGAPGRGPPGSGAGAFKQGLNGLPPIGPPGRRGTGAGPV 1053
Db 994 GPPGPPGSGPQSGASGAPGRGPPGSGAGAFKQGLNGLPPIGPPGRRGTGAGPV 1053
QY 604 GDLGAPGSPGAG 616
Db 1054 GPPGPPGPPGPPG 1066

RESULT 6
US-09-585-887-9
; Sequence 9, Application US/09585887
; Patent No. 6413742
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/585,887
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/289,578
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-585-887-9

Query Match 63.8%; Score 2224.5; DB 4; Length 1461;
Best Local Similarity 59.9%; Pred. No. 8e-133;
Matches 439; Conservative 23; Mismatches 154; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGPPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 48
Db 455 GPPGPPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 514
QY 49 GPKGSPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 108
Db 515 GPKGSPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 574
QY 109 GVMGPPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 168
Db 575 GVMGPPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 634
QY 169 GSPGPPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 212
Db 635 GSPGPPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 694
QY 213 -----TGLPFPGERGSGRGGPAGDVGAPKGPAGRRGSPGA 258
Db 695 GAMPAGNDGAKGDAGAPGAPGAGPAGDVGAPKGPAGRRGSPGA 754

QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGSPGDPGKTGPPGAGDQGRP 300
Db 755 KDGQVGRGLTGTGIPGPPGAPAGDKEGSPGSPGAGPTGARGAPGDRGEFGPPGPPGAGFAGPP 814
QY 301 GPPGPPGARGQ-----ACVMGFPFGPKGAAGPFGKAGRGVPGPPGA-----341
Db 815 GADGQPGAKGPPGDAGAKGDAGPPGAPGPPGPIGNVGAAPGAKGARGSGAPPGATGPP 874
QY 342 -----VGPAGKDGAGACGPPGAPG-----AGERGEQSGAGSPGQGLGPPGAPGPEACKP 393
Db 875 GAAGRVGPPGSGNAGPPGPPGPKGKGKGRGETGAPRPPGVEVPPGPPGPPGAGEKSP 934
QY 394 GQGVGPPDLGAPGSPGAPGEPGPTGLPGPPGRRGSGRGGPAGDVGAPKGPAGRRGSP 453
Db 935 GADGAPAGPTGPGQIAGQGVVGLPGQSGRGGPGLPGPSGEPKQSGASGERGPP 994
QY 454 GPAGP---KSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGDGRPPGPP 510
Db 995 GPMGPPGLAGPPGREGAGPAGAGSGSPGRCGSGAKGDRGETGAPGPGXGXGAPGPV 1054
QY 511 GARGQAGVMGPPGPKGAAGPFGKAGRGVPGPPGAVGPAKDGEB-----AGAQ 558
Db 1055 GPAGKSDRGRTGTPAGPAGVPGAGARGPAGPQCPRGDKGTGEGQDGRGKIHGRGFSGLQ 1114
QY 559 GPPGP-----AGPAGERGQSPAGP---GFCGLPAGPAGPGEAGKGPGEQGV 603
Db 1115 GPPGPPGSGPQSGASGAPGRGPPGSGAGAFKQGLNGLPPIGPPGRRGTGAGPV 1174
QY 604 GDLGAPGSPGAG 616
Db 1175 GPPGPPGPPGPPG 1187

RESULT 7
US-09-289-578-9
; Sequence 9, Application US/09289578
; Patent No. 6428978
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/289,578
; CURRENT FILING DATE: 1999-04-10
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-289-578-9

Query Match 63.8%; Score 2224.5; DB 4; Length 1461;
Best Local Similarity 59.9%; Pred. No. 8e-133;
Matches 439; Conservative 23; Mismatches 154; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGPPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 48
Db 455 GPPGPPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 514
QY 49 GPKGSPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 108
Db 515 GPKGSPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 574
QY 109 GVMGPPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 168
Db 575 GVMGPPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 634
QY 169 GSPGPPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 212
Db 635 GSPGPPGRRGSGRGGPAGDVGAPKGPAGRRGSPGA 694
QY 213 -----TGLPFPGERGSGRGGPAGDVGAPKGPAGRRGSPGA 258
Db 695 GAMPAGNDGAKGDAGAPGAPGAGPAGDVGAPKGPAGRRGSPGA 754


```
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 49
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-49

Query Match      62.8%; Score 2191; DB 3; Length 822;
Best Local Similarity 56.9%; Pred. No. 6.3e-131;
Matches 436; Conservative 22; Mismatches 158; Indels 150; Gaps 13;

QY 1 GPPGEPGPTGLPDPGE-----RGPGSRGFPAGDGVAGPKGPGAGRGSPGPA 48
Db 12 GPPGAPGPGQFPGPFGBFPGSGPMGPRGPPGPKNGDGEAGKPGRFGPPGPQ 71
QY 49 GPKGSPGAGRP-----GEAGLPKAGLGTGSPGSPGPDGKTGTPGPGAGOD 93
Db 72 GARGLPTAGLPGMKHGRGSLDGAKDAGAPGPKGPGSPGNGAPGQVGRGLPGR 131
QY 94 GRPGPPPPGARGQAVMGVFPKGAAGE-----PKAGERGVPPGPAVAGPKDGEA 147
Db 132 GRPGPGGTAGARGNDGAVGAAGPPTGPTGPPGPPGAVGAKGAGPQAGRGSPQGV 191
QY 148 GAQPPPGPAGPAGRGQGP-----AGSPGFQGLPQGPAGPPGAGKGEQ 192
Db 192 GEPGPPGAGAGPAGNPGADGQGKANGAPGAGPFGARGSPGQSGSPGPK 251
QY 193 GVPDGLGAP-----GPSGPA-----GPPGPTGLPFPFGRGGP 225
Db 252 GNSGEPAGPKNGDTAKGEPGATGVQPPGPGAGECKRGARGEPGSLGPPGREGP 311
QY 226 GSRFPAGDGVAGPKGAGRGSPGAPGPKGSPGAEAGLPKAGLGTGSPGSPGD 285
Db 312 GSRFPAGDGVAGPKGSPGSPGAPGPKGSPGAEAGLPKAGLGTGSPGSPGD 371
QY 286 KGTGPPGAGDGRPPGPPGARGQAVMGVFPKGAAGEPAGKAGRGVPPGPAVGA 345
Db 372 KGTGPPGAGDGRPPGAGPPGARGQAVMGVFPKGTAGEPKAGRGVPPGPAVGA 431
QY 346 CKDGEAGAQGPPGAPGAGERGEGQAGSPGFGQLPQAPGPPGAGKPGQGVPGDLGAP 405
Db 432 CKDGEAGAQGAPGAPGAGERGEGQAGSPGFGQLPQAPGPPGAGKPGQGVPGDLGAP 491
QY 406 GPSQPAGE-----PGP-----TGLPGPGRGSPGSRGFP 435
Db 492 GPSGARGRGFPGRGVGPPGAPGRNNGAPNDGAKGDTGAPGAPGQAGPLGQMP 551
QY 436 GADGVAGPKGAPGARGSPGAPKSPGCE-----AGRPGAGLFGKAGLGTGSPGSP 486
Db 552 GERGAAGLPGKGRDGDAGPKGADGSPKGDGARGLTGPIGPPGAPGAPGKAGRGSPG 611
QY 487 GPD-----GKTGPPGAPGQDGRGPPGPPGARGQ-----AGVNGFP 522
Db 612 GPTGARGAPGRGAGPPGAPGAPGADGQPGKAGGEPDGTGKGDAGPPGAPGAPG 671
QY 523 GPKGAAGFPKAGRGVPPGPPCA-----VGPAGKDGAGAQGPPGAPG-----AGER 570
Db 672 CFIGVAGPAPKGRFAGAGPAGATGFCGAAAGRVPPGPPGPNAGPPGPPGKGGKPR 731
QY 571 GEQGPAGSPGFGGLPAGPAGPGEAGKPGEQVQVFGDLGAPGSPGAP 616
Db 732 GETGAPRGVGVPPGPPGAGEXGSPGADGAPGASPTGPPQGIAG 777
```

```
RESULT 10
US-08-468-996-10
; Sequence 10, Application US/08468996
; Patent No. 6645504
; GENERAL INFORMATION:
; APPLICANT: Weiner, Howard
; APPLICANT: Miller, Ariel
; APPLICANT: Zheng, Zheng
; APPLICANT: Ahmad, Al-Sabbagh
; TITLE OF INVENTION: BYSTANDER SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION
; TITLE OF INVENTION: GLUCAGON
; FILE REFERENCE: 1010/16959-US3
; CURRENT APPLICATION NUMBER: US/08/468,996
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: US 07/843,752
; PRIOR FILING DATE: 1992-02-28
; PRIOR APPLICATION NUMBER: US 07/460,852
; PRIOR FILING DATE: 1990-02-21
; PRIOR APPLICATION NUMBER: US 07/596,936
; PRIOR FILING DATE: 1990-10-15
; PRIOR APPLICATION NUMBER: US 07/065,734
; PRIOR FILING DATE: 1987-06-24
; PRIOR APPLICATION NUMBER: US 07/454,486
; PRIOR FILING DATE: 1989-12-20
; PRIOR APPLICATION NUMBER: US 07/487,732
; PRIOR FILING DATE: 1990-03-02
; PRIOR APPLICATION NUMBER: US 07/551,632
; PRIOR FILING DATE: 1990-07-10
; PRIOR APPLICATION NUMBER: US 07/379,778
; PRIOR FILING DATE: 1989-07-14
; PRIOR APPLICATION NUMBER: US 07/607,826
; PRIOR FILING DATE: 1990-10-31
; PRIOR APPLICATION NUMBER: US 07/595,468
; PRIOR FILING DATE: 1990-10-10
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 1017
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-468-996-10
```

```
Query Match      56.0%; Score 1955; DB 4; Length 1017;
Best Local Similarity 52.5%; Pred. No. 5.4e-116;
Matches 396; Conservative 38; Mismatches 182; Indels 138; Gaps 15;

QY 1 GPPGEPGPTGLPDPGE-----RGPGSRGFPAGDGVAGPKGPGAGRGSPGPA 60
Db 25 GPPGEPGPTGLPDPGE-----RGPGSRGFPAGDGVAGPKGPGAGRGSPGPA 84
QY 61 -----GEAGLPKAGLGTGSP-----GSPGP-----DGKTPPPGPA 90
Db 85 GVKHGRGYPGLDGAKGAGAGVKGESGSPGNGSGFGPMGPRGLPGERGRTGPAAGAR 144
QY 91 GQDGRGPPGPGARGQAGVGMFP-----GPKGAGEBPKAGRGVPPGPAVGA 141
Db 145 GNDQGPAGPAGPPGVPAGGPPGAPGAKGAGTGPGRGPPDQAGTGLGPKGTGKPGIAGPK 204
QY 142 -----GKDGEAGAQGPPGAPGAGERGEGQAGSPGFGQLPQAPGPPGAGKPGEQGV 195
Db 205 GASCHNPTGDIFGAKGSAGAPGAGPFGPRGPPDQAGTGLGPKGTGKPGIAGPK 264
QY 196 GD---LGAPGSPGAPGEPGPTGLP-----GPPGREGGPGSRGFPAGDVA 237
Db 265 GEQGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 324
QY 238 GPKGAPGARGSPGAPGPKGSPGAEAGLPKAGLGTGSPGSPGPDGKTGTPGPGAGOD 297
Db 325 GPKGAPGARGSPGAPGPKGSPGAEAGLPKAGLGTGSPGSPGPDGKTGTPGPGAGOD 384
QY 298 GRPFPFPFGARGQAGVGMFPGPKGAAGBPKAGRGVPPGPAVAGPKGAGEAGAQGPP 357
```


PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/187,319
FILING DATE: 21-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN -ALPHA 1 (II)
US-08-963-825-20

Query Match 55.7%; Score 1943; DB 3; Length 1418;
Best Local Similarity 52.1%; Pred. No. 4e-115;
Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGEPGPTGLPFGPGRGSGRFFPGADGVAGPKGPAGRGSPGPKGSGEAGRP 60
DB 156 GNPGEPEGVSGPMGRGPPGPKGPDGDEAGKPKGKAGRGPPGQARGPPTGLP 215
QY 61 -----GEAGLPKAGLTGSP-----GSPGP-----DGKTGPPGPA--- 90
DB 216 GVKHRYVPLDGAKEAGAPGVKSGSGFNGSPGMPGRLPGERGRITGPAAGAR 275
QY 91 GQDGRPGPPGARGAGVGMFF-----GPKGAAGEPKKAGRGVPPGPAVGPA 141
DB 276 GNDQGPAGPPGVPGAGGPPGPAKGAAGEGFTGARGPEGAQGRGPPGSPGPA 335
QY 142 -----GKDEAGAGPPGAPGAGRGEGGAGSPGFGPLGPPAGPPGAGRGEGQVP 195
DB 336 GASGNPDTGIPGAKGSAGAPGAGAPFGPPGPPDPQATGELGPKGTGKXPIAGFK 395
QY 196 GD---LGAPGSPGAPGPTGLP-----GPPRGSGRGPFGADGVA 237
DB 396 GEQPKGEFGAGPQAGAPGAPGAGSGKRGARPGVGPFGPPGERGAPGNGRPPQDGLA 455
QY 238 GPKGAPGERSGAPGAPGKSGPGRGPRGEGAGLPKAGLTGSPGSPGPDGTPGPPAGQD 297
DB 456 GPKGAPGERSGLAGPKGANGDPGPPGEPGLPGARGLTGSPGPDGAPGQGVGSPGAFED 515
QY 298 GRPDPGPPGARGAGVGMFFPKGAAGEPKKAGRGVPPGPAVGPAKDGAGAGQPP 357
DB 516 GRPDPGPPGARGAGVGMFFPKGANGEPKAGKELPLGAPGLRGLPGKDGRTGAGPP 575
QY 358 GPAPGAPGERSGAPGSPGFGPLGPPAGPPGEGAGKPGEGQGVGPDGAPGSPGAGE---- 413
DB 576 GPAPGAPGERSGAPGSPGFGPLGPPAGPPGEGGKPGDQGVGPAAGAPGLVGRGERGFP 635
QY 414 -----PGTGLP-----GPPRGSGRGPFGADGVAGPKGPA 447
DB 636 GERGSPGAGLQGRGLPGTGTGTDGPKGASGAPGAPGAGPGLQGMPPGERGAAGIAGPK 695
QY 448 GERGS-----PGPAPGKSPGAGRPFGAGLPKAGLTGSP 483
DB 696 GDRGVGKEGEGAGKDGKGRGLTGPFGPPGAGANGKEGVGPPGAGSAGAGAPGER 755
QY 484 GSPGPDGKTGPPGAPGQGRGPPGPPGARGQAQGVMPGPKGAAGP-----GKA 534
DB 756 GETGPPGTSGIAGPPGADGQPCAKGEGQAGKGDAGAPGQGSAGFPGQGTGVTGPK 815
QY 535 GERGVPPGPA-----VGPAGKDGAGAGQGPAPA---GPAGRGEGGAGSPGFQ 582

DB 816 GARGAQPPGATGFPGAAGRVGPPGSGNGNPPGPPGSGKDGKARGDSDSPGRAGEP 875
QY 583 GLPGFAGPPGAGKPGEGQGVGPDGLGAPGPPGPA 616
DB 876 GLQGFAGPPGEGKEGPPGDDGPGSAGPPGPPQGLAG 909

RESULT 13
US-09-010-999-1
Sequence 1, Application US/09010999
Patent No. 6132976
GENERAL INFORMATION:
APPLICANT: Poole, Anthony R.
APPLICANT: Hollander, Anthony P.
APPLICANT: Billingham, R.C.
TITLE OF INVENTION: IMMUNOSAYS FOR THE MEASUREMENT OF
TITLE OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSER: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010.999
FILING DATE: 22-JAN-1998
CLASSIFICATION: 4335
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/448,501
FILING DATE: 17-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/984,123
FILING DATE: 04-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 032931/0212
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1418 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Human Type II Collagen
US-09-010-999-1

Query Match 55.7%; Score 1943; DB 3; Length 1418;
Best Local Similarity 52.1%; Pred. No. 4e-115;
Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGEPGPTGLPFGPGRGSGRFFPGADGVAGPKGPAGRGSPGPKGSGEAGRP 60
DB 156 GNPGEPEGVSGPMGRGPPGPKGPDGDEAGKPKGKAGRGPPGQARGPPTGLP 215
QY 61 -----GEAGLPKAGLTGSP-----GSPGP-----DGKTGPPGPA--- 90
DB 216 GVKHRYVPLDGAKEAGAPGVKSGSGFNGSPGMPGRLPGERGRITGPAAGAR 275
QY 91 GQDGRPGPPGARGAGVGMFF-----GPKGAAGEPKKAGRGVPPGPAVGPA 141

QY 583 GLPFPAGPGEAGKPGEGVQVGLGAPGSPGAG 616
 Db 876 GLQGPAGPGEKGEKPGDPSGAEPPGQGLAG 909

RESULT 15
 US-09-570-573-20
 ; Sequence 20, Application US/09570573
 ; Patent No. 6342361
 ; GENERAL INFORMATION:
 ; APPLICANT: Qvist, Per
 ; APPLICANT: Bonde, Martin
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 ; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/570,573
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/187,319
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gogoris, Adda C
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 4305/08701
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-527-7700
 ; TELEFAX: 212-753-6237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 20:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1418 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; IMMEDIATE SOURCE:
 ; CLONE: COLLAGEN -ALPHA 1 (II)

Search completed: October 18, 2004, 13:43:55
 Job time : 29.4569 secs

Query Match 55.7%; Score 1943; DB 3; Length 1418;
 Best Local Similarity 52.1%; Pred. No. 4e-115;
 Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGEPGTGLPDPGPPGEGGSGRGFPDADGVAGKPGAGERSGPPGAPGKSGSPGAGRP 60
 Db 156 GNPGEPPGPGVGMPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 215

QY 61 -----GEAGLPKAGLGTGSP---GSPGP-----DGKGTGPPGPA--- 90
 Db 216 GVKGHRGYPLDGAKEGAGAGVKGESGSPGPMGPRGLPGERGRITGPAGAAGAR 275

QY 91 GQDGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 141
 Db 276 GNDGQPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 335

QY 142 -----GKDGAGAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 195

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:42:06 ; Search time 79.3776 Seconds
(without alignments)
2513.149 Million cell updates/sec

Title: US-10-658-989A-2
 Perfect score: 3488
 Sequence: 1 GPPGPGPTGLPSPGRRG.....GEQVPGDLGAPSPGPAG 617

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1360919 seqs, 323318874 residues
Total number of hits satisfying chosen parameters: 1360919

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 and

```

Database : Published Applications_AA.*
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/2/pubppaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/2/pubppaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/2/pubppaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/2/pubppaa/US10D_PUBCOMB.pcp.*
17: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pcp.*
19: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pcp.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	2237.5	64.1	1057	15	US-10-104-889-20	Sequence 20, Appl
2	2237.5	64.1	1107	15	US-10-104-889-11	Sequence 11, Appl
3	2237.5	64.1	1171	15	US-10-104-889-8	Sequence 8, Appl
4	2237.5	64.1	1388	15	US-10-104-889-10	Sequence 10, Appl
5	2237.5	64.1	1464	14	US-10-104-889-21	Sequence 21, Appl
6	2237.5	64.1	1464	16	US-10-357-851-1	Sequence 1, Appl
7	2237.5	64.1	1464	16	US-10-358-024-1	Sequence 1, Appl
8	2237.5	64.1	1464	17	US-10-788-792-150	Sequence 150, App
9	2234	64.0	1449	15	US-10-402-089-8	Sequence 8, Appl
10	2234	64.0	1449	15	US-10-402-0728-8	Sequence 8, Appl
11	2233.5	64.0	1461	16	US-10-468-091-25	Sequence 25, Appl
12	2233.5	64.0	1461	14	US-10-291-265-243	Sequence 243, App
13	2231.5	63.9	1057	15	US-10-104-889-16	Sequence 16, Appl
14	2228.5	63.9	1341	14	US-10-058-124-18	Sequence 18, Appl

15	2227.5	63.9	1464	10	US-09-918-715-261	Sequence 261, Appl
16	2227.5	63.9	1464	14	US-10-060-036-159	Sequence 139, Appl
17	2227.5	63.9	1464	14	US-10-171-311-36	Sequence 36, Appl
18	2227.5	63.9	1464	14	US-10-149-352-2	Sequence 2, Appl
19	2227.5	63.9	1464	14	US-10-177-933-65	Sequence 65, Appl
20	2227.5	63.9	1464	14	US-10-301-822-28	Sequence 28, Appl
21	2227.5	63.9	1464	16	US-10-734-564-79	Sequence 79, Appl
22	2227.5	63.8	1463	15	US-10-402-089-2	Sequence 2, Appl
23	2225.5	63.8	1463	15	US-10-402-072A-2	Sequence 2, Appl
24	2225.5	63.8	1463	15	US-10-104-889-6	Sequence 6, Appl
25	2225.5	63.7	1463	15	US-10-104-889-6	Sequence 6, Appl
26	2205.5	62.8	1453	16	US-10-468-091-36	Sequence 26, Appl
27	2191	62.8	1453	16	US-10-468-091-36	Sequence 49, Appl
28	1946	55.6	1017	16	US-10-342-331-49	Sequence 49, Appl
29	1943	55.7	1014	14	US-10-639-286-10	Sequence 10, Appl
30	1943	55.7	1018	14	US-10-194-414A-1	Sequence 1, Appl
31	1932	55.4	1014	14	US-10-058-134-20	Sequence 20, Appl
32	1932	55.4	1014	16	US-10-468-091-5	Sequence 5, Appl
33	1932	55.4	1014	16	US-10-194-414A-48	Sequence 48, Appl
34	1919	55.0	714	14	US-10-468-091-6	Sequence 6, Appl
35	1919	55.0	714	14	US-10-233-895-44	Sequence 44, Appl
36	1892.5	54.3	662	14	US-10-231-581-44	Sequence 44, Appl
37	1885	54.0	501	14	US-10-328-158A-15	Sequence 15, Appl
38	1882.5	54.0	595	14	US-10-233-175-33	Sequence 33, Appl
39	1882.5	54.0	595	14	US-10-232-175-17	Sequence 17, Appl
40	1875	53.8	1466	15	US-10-342-331-48	Sequence 48, Appl
41	1875	53.8	1466	15	US-10-342-331-50	Sequence 50, Appl
42	1862	53.4	1466	15	US-10-402-089-12	Sequence 12, Appl
43	1862	53.4	1466	15	US-10-402-072A-12	Sequence 12, Appl
44	1862	53.4	1466	15	US-10-403-089-4	Sequence 4, Appl
45	1862	53.4	1466	15	US-10-403-072A-4	Sequence 4, Appl
46	1862	53.4	1466	15	US-10-403-072A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-10-104-899-20
; Sequence 20, Application US/10104899
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE O'VINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEO ID NO: 20:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1057 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 20:
 US-10-104-889-20

Query Match 64.1%; Score 2237.5; DB 15; Length 1057;
 Best Local Similarity 60.2%; Pred. No. 1.9e-104;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPFPGERGGPSRFPAGDVGAGPKGAGERSGPPA 48
 DB 297 GPPGAGEGKRGARGEPGPTGLPFGGERGGPSRFPAGDVGAGPKGAGERSGPPA 356

QY 49 GPKSGPGEAGPGEAGLPGAKGLTSGSPGPDGKTGPPAGDGRPPGPPGARGQA 108
 DB 357 GPKSGPGEAGPGEAGLPGAKGLTSGSPGPDGKTGPPAGDGRPPGPPGARGQA 416

QY 109 GVMGFPFGKGAAGPFGKAGRGVPPGCAVGPAGKDGCAQAQPPGPPGAGRGEGQPA 168
 DB 417 GVMGFPFGKGAAGPFGKAGRGVPPGCAVGPAGKDGCAQAQPPGPPGAGRGEGQPA 476

QY 169 GSPFGQLGPPAGPPGAGKPGEGVPGDLGAPGSPGAGE-----PGP---- 212
 DB 477 GSPFGQLGPPAGPPGAGKPGEGVPGDLGAPGSPGAGERGPPGPPGAGPPG 536

QY 213 -----TGLPFPGERGGPSRFPAGDVGAGPKGAGERSGPPA 536
 DB 537 GAGAPGNDGAKGADGAPGAPGSCQAPGLQMPGEGGAAGLPGPKGDRGADGADGSP 596

QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGPDGKTGPPGAGDGRP 300
 DB 597 GKDVRGLTGPFGPPGAPAGPKGSGSPGPPAGPTGARGAPGDRGEPGPPGAGPP 656

QY 301 GPPGPPGARGQ-----AGVMGFPFGKGAAGPFGKAGRGVPPGPPGA---- 341
 DB 657 GADQPGAKGEPGDAGAKGDAGPPGAPGPPGPGTGNVGAAPGAKGARGSGAGPPGATGFP 716

QY 342 -----VGPAGKDGCAQAQPPGARGP-----AGEGEGGAGSPGFGQLGPPGAGPPGAGKP 393
 DB 717 GAAGRVGPPGSGNAGPPGPPGAGKEGKGPRGETGAGRGVPPGPPGPPGAGKEGSP 776

QY 394 GEGVPGDLGAPGSPGAGEPPTGLPFGPERGGPSRFPAGDVGAGPKGAGRGSP 453
 DB 777 GADGAPAGPTGPGQGLAGQGVVGLPQGRGERGFPGLPFGSPGEPKQPSGASGERGPP 836

QY 454 GPAGP---KGSPGEAGPGEAGLPGAKGLTSGSPGPDGKTGPPAGDGRPPGPPG 510
 DB 837 GPMGPPGLAGPPGSGREGAPGAGSGPRDGSFAGKDRGETGPPGPPGAPGAPGPPV 896

QY 511 GARGQAQVMGFPFGKGAAGPFGKAGRGVPPGCAVGPAGKDGCAQAQPPGPPGAGRGEGQPA 558
 DB 897 GPAGKSGDRGETGPPAGPAGPVGAGARGPAGPQGRGDKGETGEGQDRGIGKHGPFSGIQ 956

QY 559 GPPGP-----AGPAGRGEGQPPGAGP-----GFGQLGPPAGPPGAGKPGEGQVP 603
 DB 957 GPPGPPSGPGEQPSGASGAPGPPGPPGSAFCKDGLNGLPQFTGPPGPPRGTGAGPV 1016

QY 604 GDLGAPGSPGAP 616
 DB 1017 GPPGPPGPPGPPG 1029

RESULT 2

US-10-104-889-11
 ; Sequence 11, Application US/10104889
 ; Publication No. US20040086961A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GRUSKIN, ELLIOT A.
 ; BUECHTER, DOUGLAS

BROKAW, JANE
 ZHANG, GUANGHUI
 PAOLELLA, DAVID
 TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: DILWORTH & BARRESE
 STREET: 333 EARLE OVINGTON BOULEVARD
 CITY: UNIONDALE
 STATE: NY
 COUNTRY: U.S.A.
 ZIP: 11553

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/104,889
 FILING DATE: 22-Mar-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/169,768
 FILING DATE: 09-OCT-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: STEEN, JEFFREY S
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (516) 228-8484
 TELEFAX: (516) 228-8516
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1107 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 11:
 US-10-104-889-11

Query Match 64.1%; Score 2237.5; DB 15; Length 1107;
 Best Local Similarity 60.2%; Pred. No. 2e-104;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPFPGERGGPSRFPAGDVGAGPKGAGERSGPPA 48
 DB 297 GPPGAGEGKRGARGEPGPTGLPFGGERGGPSRFPAGDVGAGPKGAGERSGPPA 356

QY 49 GPKSGPGEAGPGEAGLPGAKGLTSGSPGPDGKTGPPAGDGRPPGPPGARGQA 108
 DB 357 GPKSGPGEAGPGEAGLPGAKGLTSGSPGPDGKTGPPAGDGRPPGPPGARGQA 416

QY 109 GVMGFPFGKGAAGPFGKAGRGVPPGCAVGPAGKDGCAQAQPPGPPGAGRGEGQPA 168
 DB 417 GVMGFPFGKGAAGPFGKAGRGVPPGCAVGPAGKDGCAQAQPPGPPGAGRGEGQPA 476

QY 169 GSPFGQLGPPAGPPGAGKPGEGVPGDLGAPGSPGAGE-----PGP---- 212
 DB 477 GSPFGQLGPPAGPPGAGKPGEGVPGDLGAPGSPGAGERGPPGPPGAGPPG 536

QY 213 -----TGLPFPGERGGPSRFPAGDVGAGPKGAGERSGPPA 536
 DB 537 GAGAPGNDGAKGADGAPGAPGSCQAPGLQMPGEGGAAGLPGPKGDRGADGADGSP 596

QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGPDGKTGPPGAGDGRP 300
 DB 597 GKDVRGLTGPFGPPGAPAGPKGSGSPGPPAGPTGARGAPGDRGEPGPPGAGPP 656

QY 301 GPPGPPGARGQ-----AGVMGFPFGKGAAGPFGKAGRGVPPGPPGA---- 341
 DB 657 GADQPGAKGEPGDAGAKGDAGPPGAPGPPGPGTGNVGAAPGAKGARGSGAGPPGATGFP 716

QY 342 -----VGPAGKDGCAQAQPPGARGP-----AGEGEGGAGSPGFGQLGPPGAGPPGAGKP 393
 DB 717 GAAGRVGPPGSGNAGPPGPPGAGKEGKGPRGETGAGRGVPPGPPGPPGAGKEGSP 776

QY 394 GEGVPGDLGAPGSPGAGEPPTGLPFGPERGGPSRFPAGDVGAGPKGAGRGSP 453
 DB 777 GADGAPAGPTGPGQGLAGQGVVGLPQGRGERGFPGLPFGSPGEPKQPSGASGERGPP 836

QY 454 GPAGP---KGSPGEAGPGEAGLPGAKGLTSGSPGPDGKTGPPAGDGRPPGPPG 510
 DB 837 GPMGPPGLAGPPGSGREGAPGAGSGPRDGSFAGKDRGETGPPGPPGAPGAPGPPV 896

QY 511 GARGQAQVMGFPFGKGAAGPFGKAGRGVPPGCAVGPAGKDGCAQAQPPGPPGAGRGEGQPA 558
 DB 897 GPAGKSGDRGETGPPAGPAGPVGAGARGPAGPQGRGDKGETGEGQDRGIGKHGPFSGIQ 956

QY 559 GPPGP-----AGPAGRGEGQPPGAGP-----GFGQLGPPAGPPGAGKPGEGQVP 603
 DB 957 GPPGPPSGPGEQPSGASGAPGPPGPPGSAFCKDGLNGLPQFTGPPGPPRGTGAGPV 1016

QY 604 GDLGAPGSPGAP 616
 DB 1017 GPPGPPGPPGPPG 1029

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/104,889
 FILING DATE: 22-Mar-2002
 CLASSIFICATION: <Unknown>
 PRIORITY INFORMATION:
 APPLICATION NUMBER: US/09/169,768
 FILING DATE: 09-OCT-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: STEEN, JEFFREY S
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (516) 228-8484
 TELEFAX: (516) 228-8516
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1388 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-10-104-889-10

Query Match 64.1%; Score 2237.5; DB 15; Length 1388;
 Best Local Similarity 60.2%; Pred. No. 2.3e-104;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

```

QY 1 GPP-----GPPPTGLPPPPGRRGGPSRFFPGADGVAGPKGAGRGSPGPA 48
DB 297 GPPGAGEGKRGARGPPTGLPPPPGRRGGPSRFFPGADGVAGPKGAGRGSPGPA 356
QY 49 GPKSPGAEAGRPAGLPAGKLTGSPGSPGPKTGPAGDGRPPGPPGARGQA 108
DB 357 GPKSPGAEAGRPAGLPAGKLTGSPGSPGPKTGPAGDGRPPGPPGARGQA 416
QY 109 GVMGFPKGGAAGPFGKAGRGVPPGCAVGPAGKDGAGAGAGPPGPPGARGQA 168
DB 417 GVMGFPKGGAAGPFGKAGRGVPPGCAVGPAGKDGAGAGAGPPGPPGARGQA 476
QY 169 GSPGFQGLPPAGPPGAGRGVPPGCAVGPAGKDGAGAGAGPPGPPGARGQA 212
DB 477 GSPGFQGLPPAGPPGAGRGVPPGCAVGPAGKDGAGAGAGPPGPPGARGQA 536
QY 213 -----TGLPFPGERGGPSRFFPGADGVAGPKGAGRGSPGPA 258
DB 537 GAGAPGNDGAKGDAGAGAPGAGPGLQGMFGERGAAGLPKPKDGRDAGPKGADGSP 596
QY 259 GE-----AGEPGEAGLPKAK-----GLTSPGSPGPKTGPAGDGRPP 300
DB 597 GKDGVRGLTGPFGPPGAGAPGDKGSPGSPGAGTGCARGAPDRGEPGPPGAGPAG 656
QY 301 GPPGPPGARGQ-----AGVMGFPKGGAAGPFGKAGRGVPPGPA--- 341
DB 657 GADGQPKAGEPDKAGKADAGPPGAGPAGPPIGNVGAAGKAGSAGPAGATGPP 716
QY 342 -----VGPAGKDGAEAGQPPGAGP---AGERGEGPAGSPGFGQLPGPAGP 393
DB 717 GAAGRVGPPGPPGAGPPGPPGAGKEGKGRGETGTPAGRGVPPGPPGAGKEGSP 776
QY 394 GEQGVPCDLGAPGSPGAGPPGPPGTLPPGPPGRRGGPSRFFPGADGVAGPKGAGRGSP 453
DB 777 GADGAPAGATPFPQGIAGRGVVGLPGQRGEGRFPGLPGPGEKQKQPSGASGERGPP 836
QY 454 GPAGP---KSGPGEAGRPAGLPAGKLTGSPGSPGPKTGPAGDGRPPGPPG 510
DB 837 GPMGFPGLAGPPGSPGREGAPGAGSPGRRDGGSPGAKGDRGETGTPAGPPGAPGAPGV 896
QY 511 GARGQAGVMGFPKGGAAGPFGKAGRGVPPGCAVGPAGKDGAGAGAGPPGPPG 558
DB 897 GPAGKSGDRGETGTPAGPAGVPPGAGPAGPQGGPRDGTGETGEOCDRGIKGRHFGSLG 956

```

```

QY 559 GPPGP-----AGPAGERGEGQPPAGSP---GFQGLPPGAPGPEAGKPGEGVPP 603
DB 957 GPPGPPGSPGEGQPPGASGAGPAGPPGPPGAGAPGKXGLNGLPGTIGPPGPRGTGAGPV 1016
QY 604 GDLGAPGPPGSPAG 616
DB 1017 GPPGPPGPPGPPG 1029

```

RESULT 5

US-10-216-705-21
 ; Sequence 21, Application US/10216705
 ; Publication No. US20030096973A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Meristem Therapeutics, S.A.
 ; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me
 ; FILE OF INVENTION: obtaining Such and Their Uses
 ; FILE REFERENCE: 1149-3 DIV
 ; CURRENT APPLICATION NUMBER: US/10/216,705
 ; PRIOR FILING DATE: 2002-08-09
 ; PRIOR APPLICATION NUMBER: US 09/331,347
 ; PRIOR FILING DATE: 1999-08-17
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: Patent in version 3.11
 ; SEQ ID NO 21
 ; LENGTH: 1464
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-216-705-21

Query Match 64.1%; Score 2237.5; DB 14; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 2.4e-104;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

```

QY 1 GPP-----GPPPTGLPPPPGRRGGPSRFFPGADGVAGPKGAGRGSPGPA 48
DB 458 GPPGAGEGKRGARGPPTGLPPPPGRRGGPSRFFPGADGVAGPKGAGRGSPGPA 517
QY 49 GPKSPGAEAGRPAGLPAGKLTGSPGSPGPKTGPAGDGRPPGPPGARGQA 108
DB 518 GPKSPGAEAGRPAGLPAGKLTGSPGSPGPKTGPAGDGRPPGPPGARGQA 577
QY 109 GVMGFPKGGAAGPFGKAGRGVPPGCAVGPAGKDGAGAGAGPPGPPGARGQA 168
DB 578 GVMGFPKGGAAGPFGKAGRGVPPGCAVGPAGKDGAGAGAGPPGPPGARGQA 637
QY 169 GSPGFQGLPPAGPPGAGRGVPPGCAVGPAGKDGAGAGAGPPGPPGARGQA 212
DB 638 GSPGFQGLPPAGPPGAGRGVPPGCAVGPAGKDGAGAGAGPPGPPGARGQA 697
QY 213 -----TGLPFPGERGGPSRFFPGADGVAGPKGAGRGSPGPA 258
DB 698 GAGAPGNDGAKGDAGAGAPGAGPGLQGMFGERGAAGLPKPKDGRDAGPKGADGSP 757
QY 259 GE-----AGEPGEAGLPKAK-----GLTSPGSPGPKTGPAGDGRPP 300
DB 758 GKDGVRGLTGPFGPPGAGAPGDKGSPGSPGAGTGCARGAPDRGEPGPPGAGPAG 817
QY 301 GPPGPPGARGQ-----AGVMGFPKGGAAGPFGKAGRGVPPGPA--- 341
DB 818 GADGQPKAGEPDKAGKADAGPPGAGPAGPPIGNVGAAGKAGSAGPAGATGPP 877
QY 342 -----VGPAGKDGAEAGQPPGAGP---AGERGEGPAGSPGFGQLPGPAGP 393
DB 878 GAAGRVGPPGPPGAGPPGPPGAGKEGKGRGETGTPAGRGVPPGPPGAGKEGSP 937
QY 394 GEQGVPCDLGAPGSPGAGPPGPPGTLPPGPPGRRGGPSRFFPGADGVAGPKGAGRGSP 453
DB 938 GADGAPAGATPFPQGIAGRGVVGLPGQRGEGRFPGLPGPGEKQKQPSGASGERGPP 997
QY 454 GPAGP---KSGPGEAGRPAGLPAGKLTGSPGSPGPKTGPAGDGRPPGPPG 510

```

Db 998 GPMGPPCLAGPESGREGAPGAEGSGPRDGSFGAKXDRGETGTGAPGPPGAPGAPGV 1057
Qy 511 GARGQAGVMGFFGPKGAAGBPGKAGRGVPGPPGAVGPPAGKQGE-----AGAQ 558
Db 1058 GPAGSGDRGETGTGAPGAGVPGAGAGPAGPQGRDKGETGEQDRCIKGHRGFSGLQ 1117
Qy 559 GPRGP-----AGPAGERGEOGPAGSP---GFQGLPGPAGPGEAGKPGGQGV 603
Db 1118 GPPGPPSGEQPSGNSGAPGPRGPPGSAGAPKDGKGLNGLPFIPIGPPRGTGDAGPV 1177
Qy 604 GDLGAPGSPGAP 616
Db 1178 GPPGPPGPPGPPG 1190

RESULT 6

US-10-357-851-1
; Sequence 1, Application US/10357851
; Publication No. US20040151731A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; TITLE OF INVENTION: Method and Compositions Involving
; TITLE OF INVENTION: T-Lymphocyte Reactivity with Collagen in Aortic Tissue of
; TITLE OF INVENTION: Abdominal Aortic Aneurysm Patients
; FILE REFERENCE: 13376US
; CURRENT APPLICATION NUMBER: US/10/357,851
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-357-851-1

Query Match 64.1%; Score 2237.5; DB 16; Length 1464;
Best Local Similarity 60.2%; Pred. No. 2.4e-104;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

Qy 1 GPP-----GPPGPTGLPCCPGERGSGRFPAGDGVAGPKGAGRGSPGPA 48
Db 458 GPPGAGBEGKEGARGEPGPTGLPGERGSGRFPAGDGVAGPKGAGRGSPGPA 517
Qy 49 GPKGSPGAGRPAGAGLPGAKGLTGSFGSPGPDGKTGPPAGQDGRPPGPPGARGQA 108
Db 518 GPKGSPGAGRPAGAGLPGAKGLTGSFGSPGPDGKTGPPAGQDGRPPGPPGARGQA 577
Qy 109 GVMGPPGPKGAAGBPGKAGRGVPGPPGAVGAPGAKDGEAGQPPGAPGAGERGEOGPA 168
Db 578 GVMGPPGPKGAAGBPGKAGRGVPGPPGAVGAPGAKDGEAGQPPGAPGAGERGEOGPA 637
Qy 169 GSPGFQGLPGPAGPPEAGKPGEOQVPGDLGAPGSPGAGE-----PGP----- 212
Db 638 GSPGFQGLPGPAGPPEAGKPGEOQVPGDLGAPGSPGARGRFPGERGVQGPAGPR 697
Qy 213 -----TGLPGERGSGRFPAGDGVAGPKGAGRGSPGAPGKSP 258
Db 698 GAGAPGNDGAKDAGAPGAGSOGAPLQGMGERGAAGLPGPKGDRGADGPKGADGSP 757
Qy 259 GE-----AGRPGEAGLPGAK-----GLTSPGSPGPDGKTGPPGAPGQDGRP 300
Db 758 GKDGVRGLTGPIGPPGAPAGCDKGESGSPGAPGTGARGAPDRGEPGPPGAPGAP 817
Qy 301 GPPGPPGARGQ-----AGVMGFPKGAGBPGKAGRGVPGPPGA----- 341
Db 818 GADGQFGAKGEPGDAGKDGADGPPGAPGPPGPIGNVGAAPGAKGARGSGAPGATGFP 877
Qy 342 -----VGPAGKDGAGAGQPPGAPF---AGERGEOGPAGSFGQGLPGPAGPPEAGK 393
Db 878 GAAGRVGPPGSGNAGPPGPPGAPGKKGKGRGETGAPGPEGVGPPGPPGAPGKSP 937
Qy 394 GEQGVGDLGAPGSPGAPGEPGPTGLPGERGSGRFPAGDGVAGPKGAGRGSP 453

Db 938 GADGPAGAPGTGPGQIAGQGVVGLPGQRGERRFPGLPGPSGBFGKQGPSASGERGPP 997
Qy 454 GPAGP---KGSFGGAGRGAGLPGAKGLTGSFGSPGPDGKTGPPGAPGQDGRPPGPP 510
Db 998 GPMGPPCLAGPESGREGAPGAEGSGPRDGSFGAKXDRGETGTGAPGPPGAPGAPGV 1057
Qy 511 GARGQAGVMGFFGPKGAAGBPGKAGRGVPGPPGAVGPPAGKQGE-----AGAQ 558
Db 1058 GPAGSGDRGETGTGAPGAGVPGAGAGPAGPQGRDKGETGEQDRCIKGHRGFSGLQ 1117
Qy 559 GPPGP-----AGPAGERGEOGPAGSP---GFQGLPGPAGPGEAGKPGGQGV 603
Db 1118 GPPGPPSGEQPSGASGAPGPRGPPGSAGAPKDGKGLNGLPFIPIGPPRGTGDAGPV 1177
Qy 604 GDLGAPGSPGAP 616
Db 1178 GPPGPPGPPGPPG 1190

RESULT 7

US-10-358-024-1
; Sequence 1, Application US/10358024
; Publication No. US20040151732A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; TITLE OF INVENTION: Method and Compositions Involving Blood
; TITLE OF INVENTION: T-Lymphocytes Reactivity with Collagen
; FILE REFERENCE: 13436US
; CURRENT APPLICATION NUMBER: US/10/358,024
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-358-024-1

Query Match 64.1%; Score 2237.5; DB 16; Length 1464;
Best Local Similarity 60.2%; Pred. No. 2.4e-104;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

Qy 1 GPP-----GPPGPTGLPCCPGERGSGRFPAGDGVAGPKGAGRGSPGPA 48
Db 458 GPPGAGBEGKEGARGEPGPTGLPGERGSGRFPAGDGVAGPKGAGRGSPGPA 517
Qy 49 GPKGSPGAGRPAGAGLPGAKGLTGSFGSPGPDGKTGPPGAPGQDGRPPGPPGARGQA 108
Db 518 GPKGSPGAGRPAGAGLPGAKGLTGSFGSPGPDGKTGPPGAPGQDGRPPGPPGARGQA 577
Qy 109 GVMGPPGPKGAAGBPGKAGRGVPGPPGAVGAPGAKDGEAGQPPGAPGAGERGEOGPA 168
Db 578 GVMGPPGPKGAAGBPGKAGRGVPGPPGAVGAPGAKDGEAGQPPGAPGAGERGEOGPA 637
Qy 169 GSPGFQGLPGPAGPPEAGKPGEOQVPGDLGAPGSPGAGE-----PGP----- 212
Db 638 GSPGFQGLPGPAGPPEAGKPGEOQVPGDLGAPGSPGARGRFPGERGVQGPAGPR 697
Qy 213 -----TGLPGERGSGRFPAGDGVAGPKGAGRGSPGAPGKSP 258
Db 698 GAGAPGNDGAKDAGAPGAGSOGAPLQGMGERGAAGLPGPKGDRGADGPKGADGSP 757
Qy 259 GE-----AGRPGEAGLPGAK-----GLTSPGSPGPDGKTGPPGAPGQDGRP 300
Db 758 GKDGVRGLTGPIGPPGAPAGCDKGESGSPGAPGTGARGAPDRGEPGPPGAPGAP 817
Qy 301 GPPGPPGARGQ-----AGVMGFPKGAGBPGKAGRGVPGPPGA----- 341
Db 818 GADGQFGAKGEPGDAGKDGADGPPGAPGPPGPIGNVGAAPGAKGARGSGAPGATGFP 877
Qy 342 -----VGPAGKDGAGAGQPPGAPF---AGERGEOGPAGSFGQGLPGPAGPPEAGK 393

Db 878 GAAGRVGPPGSGNAGPPGPPAGKEGKGPRGETGACGPRGVEGPPGPPAGEKGP 937
 QY 394 GEQVPGDLGAPGSGPAGERGPTGLPDPBERGGPSRPFPGADGVAGKGPAGERGSP 453
 Db 938 GADGAPAGPCTGPGQIAGQGVVGLPQQRGERFPGLPGFSGFPGKQPSGASGERGPP 997
 QY 454 GPAGP---KGSFGACRGEAGLPGAKGLTSGPSGPGDPGKTGPPGAGODGRPPGPP 510
 Db 998 GPMGPPGLAGPPGSGREGAGAGSGPGRDGSFPAKGDGETGAGPAGPAGPAGCPV 1057
 QY 511 GARGQAVGVMGPPGKGAAGEFKAGERGVPPGPAVAGPAGKDE-----AGAQ 558
 Db 1058 GPAGKSGDRGETGAGPAGPVGARGPAGPQGRGDKGTGEGQDRGIKGRHGFSLQ 1117
 QY 559 GPPGP-----AGPAGERGEGPAGSP---GFQGLPGPAGPPGAGKPGEGV 603
 Db 1118 GPPGPPGPPGPPG 1190
 QY 604 GDLGAPGSPAG 616
 Db 1178 GPPGPPGPPGPPG 1190

RESULT 8

US-10-788-792-150
 ; Sequence 150, Application US/10788792
 ; Publication No. US20040191819A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bayer Pharmaceuticals Corporation
 ; APPLICANT: Eveleigh, Despa
 ; APPLICANT: Bigwood, Douglas
 ; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
 ; FILE REFERENCE: 5152
 ; CURRENT APPLICATION NUMBER: US/10788,792
 ; CURRENT FILING DATE: 2004-02-27
 ; PRIOR APPLICATION NUMBER: US 60/450,655
 ; PRIOR FILING DATE: 2003-02-28
 ; NUMBER OF SEQ ID NOS: 254
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 150
 ; LENGTH: 1464
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-788-792-150

Query Match 64.1%; Score 2237.5; DB 17; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 2.4e-104;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;
 QY 1 GPP-----GPGPTGLPGRGGRGPGSRGFGADGVAGKGPAGERGSPGA 48
 Db 458 GPPGAGEGKRGARGPPTGLPGRGGRGPGSRGFGADGVAGKGPAGERGSPGA 517
 QY 49 GPXGSPGACRGEAGLPGAKGLTSGPSGPGDKTGPAGQDGRPPGPPGARGQA 108
 Db 518 GPXGSPGACRGEAGLPGAKGLTSGPSGPGDKTGPAGQDGRPPGPPGARGQA 577
 QY 109 GVWGFPGKGAAGEPGKAGRGVPPGAVGAPKAGEAGAGAGQPPGAPAGERGQGA 168
 Db 578 GVWGFPGKGAAGEPGKAGRGVPPGAVGAPKAGEAGAGAGQPPGAPAGERGQGA 637
 QY 169 GSPGFQGLPGPAGPPGAGKFGGQGVPGDLGAPGSPGAGE-----PGP----- 212
 Db 638 GSPGFQGLPGPAGPPGAGKFGGQGVPGDLGAPGSPGAGERGFPGERGVQPPGAPGR 697
 QY 213 -----TGLPGRGGRGPGSRGFGADGVAGKGPAGERGSPGAPKXGP 258
 Db 698 GARGPNDGAKDAGAPGAPGQAGFGLQMPGERGAAGLPGKPKDRGDAGPKGADGSP 757
 QY 259 GE-----AGPGEAGLPGAK-----GLTGSQSGSGSPGAGTGAAGACDRGEGPAGDGRP 300
 Db 758 GKDGVRLTGTPGPPGAPGDKGSGSGSPGAGTGAAGACDRGEGPAGDGRP 817

QY 301 GPPGPPGARGQ-----AGVMGFPFGKGAAGEPKKAGERGVPPGPA----- 341
 Db 818 GADGPGFAGKEGFGDAGAKGDAGPPGAPGAPGPPGIGNVGAPGAKGARGSGAPPGATGFP 877
 QY 342 -----VGPAGKDEAGAGACQPPGAPG---AGERGQGPAGSPGFGQLPGPAGPGEAGKP 393
 Db 878 GAAGRVGPPGSGNAGPPGPPAGKEGKGPRGETGACGPRGVEGPPGPPAGEKGP 937
 QY 394 GEQVPGDLGAPGSGPAGERGPTGLPDPBERGGPSRPFPGADGVAGKGPAGERGSP 453
 Db 938 GADGAPAGPCTGPGQIAGQGVVGLPQQRGERFPGLPGFSGFPGKQPSGASGERGPP 997
 QY 454 GPAGP---KGSFGACRGEAGLPGAKGLTSGPSGPGDPGKTGPPGAGODGRPPGPP 510
 Db 998 GPMGPPGLAGPPGSGREGAGAGSGPGRDGSFPAKGDGETGAGPAGPAGPAGCPV 1057
 QY 511 GARGQAVGVMGPPGKGAAGEFKAGERGVPPGPAVAGPAGKDE-----AGAQ 558
 Db 1058 GPAGKSGDRGETGAGPAGPVGARGPAGPQGRGDKGTGEGQDRGIKGRHGFSLQ 1117
 QY 559 GPPGP-----AGPAGERGEGPAGSP---GFQGLPGPAGPPGAGKPGEGV 603
 Db 1118 GPPGPPGPPGPPG 1190
 QY 604 GDLGAPGSPAG 616
 Db 1178 GPPGPPGPPGPPG 1190

RESULT 9

US-10-402-089-8
 ; Sequence 8, Application US/10402089
 ; Publication No. US20040005663A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bell, Marcum P.
 ; APPLICANT: Neff, Thomas B.
 ; APPLICANT: Polarek, James W.
 ; APPLICANT: Seeley, Todd W.
 ; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
 ; FILE REFERENCE: FP0402.3 CON
 ; CURRENT APPLICATION NUMBER: US/10/402,089
 ; CURRENT FILING DATE: 2003-03-26
 ; PRIOR APPLICATION NUMBER: US 09/709,700
 ; PRIOR FILING DATE: 2000-11-10
 ; NUMBER OF SEQ ID NOS: 72
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 8
 ; LENGTH: 1449
 ; TYPE: PRT
 ; ORGANISM: Sus scrofa
 US-10-402-089-8

Query Match 64.0%; Score 2234; DB 15; Length 1449;
 Best Local Similarity 62.0%; Pred. No. 3.6e-104;
 Matches 438; Conservative 21; Mismatches 154; Indels 94; Gaps 13;
 QY 1 GPPPEPPTGLPGRGGRGPGSRGFGADGVAGKGPAGERGSPGAPKGSFGEAGRP 60
 Db 472 GARGEPAGLPGPPGGRGPGSRGFGADGVAGKGPAGERGSPGAPKGSFGEAGRP 531
 QY 61 GEAGLPGAKGLTSGPSGPGDKTGPAGQDGRPPGPPGARGQAVWGFPGKGA 120
 Db 532 GEAGLPGAKGLTSGPSGPGDKTGPAGQDGRPPGPPGARGQAVWGFPGKGA 591
 QY 121 GEPKAGERGVPPGPAVAGPAGKDEAGAGAGQPPGAPAGERGQGPAGSPGQGLPGPA 180
 Db 592 GEPKAGERGVPPGPAVAGPAGKDEAGAGAGQPPGAPAGERGQGPAGSPGQGLPGPA 651
 QY 181 GPPGACGKPGEQGVPGDLGAPGSPGAGE-----PGP----- 212
 Db 652 GPPGACGKPGEQGVPGDLGAPGSPGAGERGFPGERGVQPPGAPGRGANGAGNDGAK 711
 QY 213 --TCLPFPGERGPGSRGFGADGVAGKGPAGERGSPGAPKGSFGEAGRP-----A 261

Db	712	GDAGAPGSPGAPGLQGVGPERGAAGLPQPKGDRDAGPKGADGAPKDGVRGLTGP	771
Qy	262	GRGEAGLPGAAGLPGSGSPGDDGKTGPPGPAQDGRPGP	317
Db	772	GPFGPAGAGDKETGTPSGPAGTGARGAPDGRGEPGPPGAFGAGPGADGQFGAKGGP	831
Qy	318	---PGPKGAAGEPGKAGERGVFPQGA	365
Db	832	TGPPGPGIGSVGAPGPKAGASAGPPGATGPPGAAGRVGPPGPSNAGPPGPPGPAKGEKS	891
Qy	366	---RGEQGPAGSPGFOGLPGAPGPPGAGKXPGEOGVFDLGAPGSPGAFGEPGTGLPGP	422
Db	892	KGRPGTGTGAPRGPAEAGPPGPPGPAKESGFGADGPAGAPGTGPGQIAGORGVVGLPGQ	951
Qy	423	PBRGPGSPGRFGADGVAGPKGPAGBRGSPGAPGKSGPGEAGRPGEAGLPGAAGLGTGS	482
Db	952	RGRGPGFLPGSPGEPKQGPSPGSGRGPFGMPG	1008
Qy	483	P---GSPGPD---GKTGPGAPGDGRGPPGPGARGQAGVMGFPKGAAGEPKAGE	536
Db	1009	PGRDGAPGPGKDRGESGAPGPPGAPGAPGAPGVGPAKSGDRGETGAPGAPGVGPVGA	1068
Qy	537	RGVPGPGVGAAGVPAKGDGE	572
Db	1069	RGAPGPGRGDKETGEQGDGRGKIKHRGFSGLQGPFGSPGCEQGPSGASGAPGPRGP	1128
Qy	573	QGPAGSP---GFOGLPGAPGPPGAGKXPGQGVPGDILGAPGSPGAPG	616
Db	1129	PGSAGAPGKQCLNGLPGFIPGPPGPRGTGAGFPVGPFGPPGPPGPPG	1175

```

RESULT 10
US-10-402-072A-8
; Sequence 8, Application US/10402072A
; Publication No. US20040018592A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seesley, Todd W.
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
; FILE REFERENCE: FPO402.2 CON
; CURRENT APPLICATION NUMBER: US/10/402,072A
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 8
; LENGTH: 1449
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-402-072A-8

```

Db	652	GPFGGAGKPGGCGQVGDLDGAPGSPGARGGRGPPGGRGVGCGPPGAPGPRGAGCAPGNDGNK	711
Qy	213	--TGLPPGPGEGRGGSPRRFPGADGVAGPKGAGERSFGPAGPKSGPGE-----A	261
Db	712	GDAGAFGAPGSGQCAPGLQGMFPERGGAAGLPGPKDGRGDAGPKGADGAPKDGVRGLTGP	771
Qy	262	GRGGEAGLPGAKLITGSPSPGPDGKTGPPGAGQDGRGPP--PCPPGARGCAGVMGF-	317
Db	772	GPPGAPAGPDGKGTGSPGAPGFTGARGAPGDRGEPFGPAGFAGPPGADGQPGAKGFP	831
Qy	318	---PGPKGAAGEPKGAGERGVDPGCA-----VGPAGKDGGEAGAGCPGPPGAPGAGE	365
Db	832	TGPPGPIGSGVAGPFGKARGSAGPPGATGPGNAGFVGPFGSGNAGPFGPPGAPGKEGS	891
Qy	366	---RGEGGAGSGPFGOGLPGPAGPPGEGAKPGEGQGVPGDLGAPGSGGAPGPGTGLPGP	422
Db	892	KGPRGTGTPACRGEAGPFGPPGAPGKEGSPGADGAPAGPTGPGQCIAGQRGVVGLPGQ	951
Qy	423	PGRGGGSGSEFGGADGVAGPKGPACERGSPPGAPKPGSPGAGRGEGAGLPGAKLITGS	482
Db	952	RGRGPPGLPGPGEKPGKQSPGSGRGPFGPMGP---PGLAGPPGSGRGAPGCAEGS	1008
Qy	483	P---GSPGPPD---GKTGTPGAPAGQDGRGPPGPPGARGGAGVMGFGPKGAAGEPGKAGE	536
Db	1009	PGRDGAPGKDRGESGAPGPGCAPGAPGPPVGPAGKSGDRGETGPAGPAGVPGPVGA	1068
Qy	537	RGVPGPPGAVGPAKPDGE-----AGAGGPPGP-----AGPAGERGE	572
Db	1069	RGAPGQPPRGDKGTGEGQDRIKGHRGFSGLQGPFGPPGSGEGCGSGAGSPGAPRGP	1128
Qy	573	QGPAGSP---GFOGLPGAPGPGGCAKPGEGQGVPGDLGAPGSGGAPG	616
Db	1129	PGSAGAPKDKGLNGLPGPIGPPGPRGTGDAGVPGPPGPPGPPG	1175
RESULT 11			
US-10-468-091-25			
; Sequence 25, Application US/10468091			
; Publication No. US20040157329A1			
; GENERAL INFORMATION:			
; APPLICANT: ADP Pharmaceutical Pty Limited			
; APPLICANT: The University of Sydney			
; TITLE OF INVENTION: Matrix gene expression in chondrogenesis			
; FILE REFERENCE: 500311			
; CURRENT APPLICATION NUMBER: US/10/468,091			
; CURRENT FILING DATE: 2003-08-13			
; PRIOR APPLICATION NUMBER: AU PR3116			
; PRIOR FILING DATE: 2001-02-15			
; NUMBER OF SEQ ID NOS: 42			
; SOFTWARE: Patent in version 3.1			
; SEQ ID NO 25			
; LENGTH: 1461			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-10-468-091-25			

Query Match	64.0%	Score 2233.5	DB 16	Length 1461
Best Local Similarity	60.0%	Pred. No. 3	8e-104	
Matches 440	Conservative	23	Mismatches 153	Indels 117
Gaps				12
QY	1	GPP-----	GPFGTGLPPPPGKGGPFGADGVAGPKPAGRGSGGPA	48
Db	455	GPPGPAEGEKGRCARGPFGTGLPPPPGKGGPFGADGVAGPKPAGRGSGGPA	514	
QY	49	GPKSGPCEAGRPGEAGLPGAKGLTGPSPGPGDKTGGPPGACQDGRPPPPGPPGARGQA	108	
Db	515	GPKSGPCEAGRPGEAGLPGAKGLTGPSPGPGDKTGGPPGACQDGRPPPPGPPGARGQA	574	
QY	109	GVMGFPKPKGAAGBPGKAGRGVFPBPAGVGPAGKDGEGAQAQPPGAPGAGRGSGGPA	168	
Db	575	GVMGFPKPKGAAGBPGKAGRGVFPBPAGVGPAGKDGEGAQAQPPGAPGAGRGSGGPA	634	
QY	169	GSPGFQGIKPGPAPPPGAGKPGEOGVPGDLCAGPSPGPAE-----	PGP----	212

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/104,889
FILING DATE: 22-Mar-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1057 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-10-104-889-16

Query Match 64.0%; Score 2231.5; DB 15; Length 1057;
Best Local Similarity 60.0%; Pred. No. 3.8e-104;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;
QY 1 GPP-----GPPGTGLPGRGGRGPGSRFPAGDVGAGPKGAGRGSPGPA 48
DB 297 GPPGAGEEGRKARGEPGTGLPGRGGRGPGSRFPAGDVGAGPKGAGRGSPGPA 356
QY 49 GPKSGPGEAGRPBAGLPGAKGLTSGPSGPDGKTGPPAGDQGRPPGPPGARGQA 108
DB 357 GPKSGPGEAGRPBAGLPGAKGLTSGPSGPDGKTGPPAGDQGRPPGPPGARGQA 416
QY 109 GVMGFPQKGAAGBPGKAGRGVPPGAVGPAKDGAGQAQPPGAGRGSGGPA 168
DB 417 GVMGFPQKGAAGBPGKAGRGVPPGAVGPAKDGAGQAQPPGAGRGSGGPA 476
QY 169 GSRGFGQLPGAPGPGAGKGEQGVGDLGAPSGPAGE-----PGP---- 212
DB 477 GSPGFQGLPGAPGPGAGKGEQGVGDLGAPSGPAGE-----PGP---- 536
QY 213 -----TGLPFPGRGGRGPGSRFPAGDVGAGPKGAGRGSPGPKGSP 258
DB 537 GANGAPGNDGAKGDAGAPGAGSGQAPQLQMPGECGAGLPGPKGRGDAGPKGADGSP 556
QY 259 GE-----AGRPGENGLPKAK-----GLTSGPSGPDGKTGPPGAGQDGRP 300
DB 597 GKDGVRLGTGPIGPPGAPAGFGDKGSGPSGPAFTGARGAPGDRGEPGPPGAGFAGPP 656
QY 301 GPPGPPGARGQ-----AGVMGFPQKGAAGBPGKAGRGVPPGPA----- 341
DB 657 GADQCPGAKGFPDAGAKGDAGFPGAPGAPGPIGNVGAFAKAGAGSAGFPATGFP 716
QY 342 -----VGPAGKGBAGAGGPPGAPG---AGERGQGPAGSPGQGLPGPAGPPGEAGKP 393
DB 717 GAACRVGPPGPGSGNAGPPGPPGACKEGKGRGTGTPAGRPGEVGPFGPPGAGRGSP 776
QY 394 GEQGVPGDLGAPGSPGAPGPTGLPGPPGGRGPGSRFPAGDVGAGPKGAGRGSP 453
DB 777 GADGAPAGPTGPPGPIAGQGVVGLPQGRGERFPGLPGPSGPPGQKGSAGSGRKP 836
QY 454 GPAGS---KGSFGEAGRGEAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRFPGPP 510
DB 837 GPMGPPGLAGFPFSGREGAFAESGSPGRDGSFGAKGDRGTGTPAGPPGAPGAPGAPV 896
QY 511 GARGQAGVMGFPFGKGAAGEGKAGRGVPPGAVGPAKDGK-----ACAQ 558
DB 897 GPAGKSGDRGTGTPAGPAGPAGVGPAGARGPAGPQGRDGTGKGTGEGQGRGKGRHGRGSLQ 956
QY 559 GPPGP-----AGPAGERGEQGPAGSP---GFGQLPGPAGPPGAGKPGEGQVP 603

DB 957 GPPGPPGSPGEGGSCASGAPGPRGPPGASAGAPKDGGLNGLPGPIGPPGPRGTCDAGPV 1016
QY 604 GDLGAPGSPGAP 616
DB 1017 GPPGPPGPPGPPG 1029
RESULT 14
US-10-058-124-18
Sequence 18, Application US/10058124
Publication No. US20030119058A1
GENERAL INFORMATION:
APPLICANT: Ovlst, Per
Bonde, Martin
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
in Body Fluids, A Test Kit and Means for Carrying Out the
Method and Use of the Method to Diagnose the Presence of
Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/058,124
FILING DATE: 29-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/570,573
FILING DATE: 2002-MAY-12
APPLICATION NUMBER: 08/187,319
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-058-124-18

Query Match 63.9%; Score 2228.5; DB 14; Length 1341;
Best Local Similarity 60.0%; Pred. No. 6.4e-104;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;
QY 1 GPP-----GEPGTGLPGRGGRGPGSRFPAGDVGAGPKGAGRGSPGPA 48
DB 334 GPPGAGEGKARGEPGTGLPGRGGRGPGSRFPAGDVGAGPKGAGRGSPGPA 393
QY 49 GPKSGPGEAGRGEAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 394 GPKSGPGEAGRGEAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQA 453

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:31:10 ; Search time 24.4569 Seconds
(without alignments)
2427.363 Million cell updates/sec

Title: US-10-658-989A-2

Perfect score: 3488
Sequence: 1 GPPGEGPTGLPFGPGRGG.....GEGVFGDLAGPSPGAGG 617

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2228.5	63.9	1464	1 CGHUIS	collagen alpha 1(I)
2	2205.5	63.2	1453	2 S21626	collagen alpha 1(I)
3	2151.5	61.7	1042	1 CGCHIS	collagen alpha 1(I)
4	2011.5	57.7	671	1 CGRTIS	collagen alpha 1(I)
5	1944	55.7	1418	2 T45467	collagen alpha 1(I)
6	1943	55.7	1497	1 CGHUC	collagen alpha 1(I)
7	1932	55.4	1419	2 A41182	collagen alpha 1(I)
8	1932	55.4	1487	2 B41182	collagen alpha 1(I)
9	1898	54.4	779	1 CGBOIS	collagen alpha 1(I)
10	1890.5	54.2	1486	1 B40333	collagen alpha 1(I)
11	1883	54.0	1492	2 A40333	collagen alpha 1(I)
12	1857.5	53.3	1464	2 S58556	collagen alpha 1(I)
13	1856	53.2	1049	1 CGBO7S	collagen alpha 1(I)
14	1853	53.1	1466	1 CGHUVL	collagen alpha 1(I)
15	1840	52.8	886	2 I50694	collagen alpha 1(I)
16	1822.5	52.3	1496	1 CGHUV	collagen alpha 2(V)
17	1804	51.7	1497	2 T49607	procollagen type V
18	1788.5	51.3	1373	1 A43291	collagen alpha 2(I)
19	1764	50.6	1366	1 CGHUVS	collagen alpha 2(I)
20	1702.5	48.8	1838	1 CGHUVI	collagen alpha 1(V)
21	1689.5	48.4	1843	2 S18803	collagen alpha 1(V)
22	1644	47.1	1806	1 CGHUIE	collagen alpha 1(X)
23	1640.5	47.0	1414	1 S23809	collagen alpha 2(I)
24	1612	46.2	1027	2 S28774	collagen alpha cha
25	1587	45.5	1024	2 S18251	collagen alpha 1(X)
26	1585	45.4	1546	1 CGHUVZ	collagen alpha 2(X)
27	1584	45.4	1691	1 S22917	collagen alpha 5(I)
28	1583.5	45.4	673	1 CGBO6C	collagen alpha 1(I)
29	1568	45.0	2944	2 A54849	collagen alpha 1(V)

30 1556.5 44.6 1549 2 148103 type VII collagen
31 1550.5 44.5 888 2 S28791 collagen alpha 1(X)
32 1547 44.4 1690 1 CGHUIB collagen alpha 4(I)
33 1530 43.9 1763 2 S16366 collagen alpha 2(I)
34 1526 43.8 964 1 CGCH2S collagen alpha 2(I)
35 1522 43.6 1669 1 CGHUF collagen alpha 1(I)
36 1510 43.3 920 2 A45748 collagen alpha 1(V)
37 1508 43.2 1603 2 S23810 collagen alpha 1(X)
38 1503.5 43.1 1669 1 CGMS4B collagen alpha 1(I)
39 1489.5 42.7 1670 1 CGHUVB collagen alpha 3(I)
40 1481 42.5 1712 1 CGHUV2 collagen alpha 2(I)
41 1471.5 42.2 812 2 S31521 collagen COLF1 - f
42 1463 41.9 1759 2 T29351 collagen alpha 2(I)
43 1459.5 41.8 1758 2 T29350 hypothetical prote
44 1456 41.7 1142 2 JX0369 collagen alpha 1(X)
45 1437.5 41.2 1752 2 A45407 collagen alpha 3(I)

ALIGNMENTS

RESULT 1

CGHUIS

collagen alpha 1(I) chain precursor - human

N/Alternate names: procollagen alpha 1(I) chain

C/Species: Homo sapiens (man)

C/Date: 12-Aug-1981 #sequence revision 04-Oct-1996 #text change 09-Jul-2004

C/Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S1:

5269; A29439; I53466; A02852; I37247

R/D/Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, P.; Pretorius, P.J.

Gene 67, 105-115, 1988

A/Title: Complete nucleotide sequence of the region encompassing the first twenty-five

A/Reference number: I60114; MUID:88329734; PMID:2843432

A/Accession: I60114

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-369, 'L', 371-589 <DAL>

A/Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNI

R/Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Proc

Biochem. J. 253, 919-922, 1988

A/Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human

A/Reference number: S01143; MUID:89025644; PMID:3178743

A/Accession: S01143

A/Molecule type: mRNA

A/Residues: 1-472 <TRO>

A/Cross-references: EMBL:X07884; NID:G30015; PIDN:CAA30731.1; PID:G30016; GB:M36546; NI

A/Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988

R/Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.

Nature 310, 337-340, 1984

A/Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation o

A/Reference number: A93335; MUID:84270697; PMID:6462220

A/Accession: A93335

A/Molecule type: DNA

A/Residues: 1-58, 'Q', 60-181 <CHU>

A/Cross-references: EMBL:X00820; NID:G35657; PIDN:CAA25394.1; PID:G35658

R/Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.

J. Biol. Chem. 262, 15151-15157, 1987

A/Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene en

A/Reference number: I55254; MUID:88033098; PMID:2822714

A/Accession: I55254

A/Status: translation not shown; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-45 <ROS>

A/Cross-references: GB:J02829; NID:G180387; PIDN:AAA51993.1; PID:G180388

R/Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gellinas, R.E.

Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987

A/Title: Regulatory elements in the first intron contribute to transcriptional control

A/Reference number: A39943; MUID:88097389; PMID:3480516

A/Accession: A39943

A/Molecule type: DNA

A/Residues: 1-34 <BOR>

A/Cross-references: GB:J03559; NID:G180876; PIDN:AAA52052.1; PID:G553238

R/Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

- J. Biol. Chem. 260, 2315-2320, 1985
 A>Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
 A'Reference number: 155237; MUID:85130970; PMID:2857713
 A'Accession: 155237
 A>Status: translation not shown; translated from GB/EMBL/DBJ
 A'Molecule type: mRNA
 A'Residues: 1-34 <CH2>
 A'CROSS-references: GB:M10627; NID:g180383; PIDN:AAA51992.1; PID:G553226
 R'Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollish
 J. Biol. Chem. 265, 6312-6317, 1990
 A>Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina
 some, type VII.
 A'Reference number: A35233; MUID:90202908; PMID:2318855
 A'Accession: A35233
 A'Molecule type: protein
 A'Residues: 33-52 <WIR>
 A>Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
 R'Veil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
 EMBO J. 8, 1705-1710, 1989
 A>Title: A base substitution in the exon of a collagen gene causes alternative splicing
 A'Reference number: S09400; MUID:89356643; PMID:2767050
 A'Accession: S09400
 A'Molecule type: mRNA
 A'Residues: 156-183 <WEI>
 R'Click, E.M.; Bornstein, P.
 Biochemistry 9, 4599-4706, 1970
 A>Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1
 A'Reference number: A90567; MUID:71038625; PMID:5529814
 A'Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
 A'Accession: B90567
 A'Molecule type: protein
 A'Residues: 162-198, 'Z', '200-201, 'Z', '203-206, 'Z', '208-209, 'Z', '211-228, 'B', '230, 'BB', '233, 'Z'
 A'Experimental source: skin
 A>Note: evidence for 170-allysine
 R'Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, F.
 Eur. J. Biochem. 192, 153-159, 1990
 A>Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
 A'Reference number: S11372; MUID:90382436; PMID:2169412
 A'Accession: S11372
 A'Molecule type: protein
 A'Residues: 175-187, '274-287, 'P', '289 <BAE>
 A>Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion
 R'Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
 J. Biol. Chem. 266, 21827-21832, 1991
 A>Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
 cooperative melting of intact type I collagen.
 A'Reference number: 155342; MUID:92042092; PMID:1718994
 A'Accession: 155342
 A>Status: translated from GB/EMBL/DBJ
 A'Molecule type: mRNA
 A'Residues: 258-268, '1347-1357 <DEA>
 A'CROSS-references: GB:S67495; NID:g239007; PIDN:AB20350.1; PID:G239008
 A>Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
 R'Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
 J. Biol. Chem. 245, 5042-5048, 1970
 A>Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
 A'Reference number: A92069; MUID:71001508; PMID:4319110
 A'Accession: A92069
 A'Molecule type: protein
 A'Residues: 263-268 <MOR>
 A'Experimental source: skin
 A>Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
 R'Labhard, M.E.; Hollister, D.W.
 Matrix 10, 124-130, 1990
 A>Title: Segmental amplification of the entire helical and telopeptide regions of the co
 A'Reference number: S15989; MUID:90326017; PMID:2374517
 A'Accession: S15989
 A'Molecule type: mRNA
 A'Residues: 281-302, '402-420, '823-843, '925-944, '1026-1045, '1143-1162 <LAB>
 R'Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N.
 Connect. Tissue Res. 29, 1-11, 1993
 A>Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
 A'Reference number: 152505; MUID:93339042; PMID:8339541
 A'Accession: 152505
- A'Accession: 152905
 A>Status: translated from GB/EMBL/DBJ
 A'Molecule type: mRNA
 A'Residues: 342-352, 'C', '354-359 <W12>
 A'CROSS-references: GB:S64717; NID:g408195; PIDN:AB27677.1; PID:g408196
 A>Note: mutant sequence from patient with osteogenesis imperfecta
 R'Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
 Biochemistry 22, 5213-5223, 1983
 A>Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha1
 A'Reference number: A90476; MUID:84080385; PMID:6689127
 A'Accession: A90476
 A'Molecule type: mRNA
 A'Residues: 425-1250, 'X', '1252-1328, 'S', '1330-1390, 'X', '1392-1464 <BER>
 A'CROSS-references: GB:K01228; NID:g180391; PIDN:AAA51995.1; PID:g180392
 A>Note: sequence partially completed for missing nucleotides by A29439
 R'Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
 J. Biol. Chem. 260, 691-694, 1985
 A>Title: Multixon deletion in an osteogenesis imperfecta variant with increased type II
 A'Reference number: A22161; MUID:85104934; PMID:2981843
 A'Accession: A22161
 A'Molecule type: DNA
 A'Residues: 472-594, 'R', '596-607 <CH3>
 A'CROSS-references: GB:K03178; GB:K03179; NID:g179612; NID:g179613; PIDN:AAA51847.1; PID
 A>Note: the authors translated the codon CGT for residue 595 as Pro
 R'Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
 Am. J. Hum. Genet. 45, 1034-1040, 1990
 A>Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained
 A'Reference number: A35336; MUID:90252792; PMID:2339700
 A'Accession: A35336
 A'Molecule type: mRNA
 A'Residues: 710-720, 'B', '722-737, 'E', '739-745 <WAL>
 A>Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
 R'Forlino, A.; Zollezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes
 Hum. Mol. Genet. 3, 2201-2206, 1994
 A>Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the c
 A'Reference number: 154365; MUID:95187161; PMID:7881420
 A'Accession: 154365
 A>Status: translated from GB/EMBL/DBJ
 A'Molecule type: DNA
 A'Residues: 746-766, 'S', '768-781 <FOR>
 A'CROSS-references: GB:L47667; NID:g1009093; PIDN:AB59576.1; PID:g1009094
 R'Chessier, S.D.; Wallis, G.A.; Byers, P.H.
 J. Biol. Chem. 268, 18218-18225, 1993
 A>Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
 A'Reference number: A47426; MUID:93352646; PMID:8349697
 A'Accession: A47426
 A'Molecule type: mRNA
 A'Residues: 1179-1276, 'H', '1278-1336, '1339-1387, 'R', '1389-1464 <CHE>
 A'CROSS-references: GB:S64596; NID:g407589; PIDN:AB27856.1; PID:g407590
 A>Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBIPI:136445)
 A>Note: does not represent an experimentally determined sequence but three different mut
 A'Accession: B47426
 A'Molecule type: mRNA
 A'Residues: 1179-1464 <CH4>
 A'Experimental source: normal dermal fibroblast culture
 A'Accession: C47426
 A'Molecule type: mRNA
 A'Residues: 1179-1276, 'H', '1278-1464 <CH5>
 A'Experimental source: fetal cell 86-237
 A'Accession: D47426
 A'Molecule type: mRNA
 A'Residues: 1179-1336, '1339-1464 <CH6>
 A'Experimental source: fetal cell 86-146
 A'Accession: E47426
 A'Molecule type: mRNA
 A'Residues: 1179-1387, 'R', '1389-1464 <CH7>
 A'Experimental source: fetal cell 88-251
 R'Cohn, D.H.; Apone, S.; Eyle, D.P.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Nic
 J. Biol. Chem. 263, 14605-14607, 1988
 A>Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptide c
 A'Reference number: 155269; MUID:89008319; PMID:3170557
 A'Accession: 155269

A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1187-1194, 'C', 1196-1220 <COH>
A;Cross-references: GB:M23213; NID:G340842; PIDN:AAB59363.1; PID:G499622
A;Note: mutant sequence from a patient with mild osteogenesis imperfecta
R;Maekelae, J.K.; Raassina, M.; Virta, A.; Vuorio, E.
Nucleic Acids Res. 15, 349, 1988
A;Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

Query Match 63.9%; Score 2228.5; DB 1; Length 1464;
Best Local Similarity 60.0%; Pred. No. 4.2e-105;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

Qy 1 GPP-----GPGGTGLGPPGSGRSGRFGADGAGVAGPKGPGAGSGGPPGA 48
Db 458 GPPGAGEGKRGARGEPGTGLPFPGERGCGSRGFPAGDAGVAGPKGPGAGSGGPPGA 517
Qy 49 GPKSGPEAGRPAGLPGAKGLTGTSGSPGPPGKTGTPFAGQDGRPGPPGPPGARGQA 108
Db 518 GPKSGPEAGRPAGLPGAKGLTGTSGSPGPPGKTGTPFAGQDGRPGPPGPPGARGQA 577
Qy 109 GVMGFPDPMAGPAGVAGRGVPPGPGAVGPGAGKDGEGAGAQPPGPGAGSGGPPGA 168
Db 578 GVMGFPDPMAGPAGVAGRGVPPGPGAVGPGAGKDGEGAGAQPPGPGAGSGGPPGA 637
Qy 169 GSPGFGGLGPPAGPPGAGKPGEGVPGDLCAPGSPGAGE-----PGP----- 212
Db 638 GSPGFGGLGPPAGPPGAGKPGEGVPGDLCAPGSPGARGERFPGERGVQGGPPGAGPR 697
Qy 213 -----TGLPGPGERGPGSRGFPAGDAGVAGPKGPGAGSGGPPGAGKGP 258
Db 698 GANGAPGNDGAKGADAGAPGAGPSGAPGLQMPGERGAAGLPKPGDRGAGPKGADGSP 757
Qy 259 GE-----AGPGEAGLPGAK-----CLTGSPPSGPDGKTGTPGPGAGQDGRP 300
Db 758 GKGVVRGLTGIPGPPGAGAPDKGESGSPGAPGTGARGAPDGRGPPGPPGAGFAGPP 817
Qy 301 GPPPPPPGARGQ-----AGVMGFPDPMAGPAGVAGPKGPGAGSGGPPGPA----- 341
Db 818 GADGQPKAGKPPGADAGKADGAPGAGPAGPGPIGNVAGPAGKARGSPGATGPP 877
Qy 342 -----VPGADKDGAGAGPPGAPG---AGEREQSPAGSPGPGGLGPPGAGPPGAGKGP 393
Db 878 GAAGRVGPPGSGNAGPPGPPGAGKGGKGRGETGAPRGPSVGPVPPGPPGAGKGP 937
Qy 394 GEQGVGDLGAPGSPGAPGPGTGLPGPDERGSGSRGPPGADGAGVAGPKGPGAGSGP 453
Db 938 CADGPAGAGTGGQGTAGQGVVGLPGQGERGFGLPGSPGPGKQSGSASGERGPP 997
Qy 454 CPAGE---XGSPGABRGEAGLPGAKGLTGSPPGPDGKTGTPGPGAGQDGRPPGPPG 510
Db 998 GPMGPPGLAGPPGSRGREGAPGABGSPGRDGSAGKADGRGTGPPGPPGAGKAGAPGV 1057
Qy 511 GARGQAGVMGPPGKAGGEPKAGRGVPPGPGAVGPKGDKG-----AGAQ 558
Db 1058 GPAGSGDRGTGPPGAPGVPVPGARGAPGQPRGDKGTGQGRGKLGHRGFGSLQ 1117
Qy 559 GPPGP-----AGPAGERGEQGPAGSP---GFGGLGPPGPPGAGKPGEGQGVF 603
Db 1118 GPPGPPSGEGSGASGAPGPPGPGSAGAPGKGLGLGPGIPGPPGRGRTGDAGPV 1177
Qy 604 GDLGAPGSGAG 616
Db 1178 GPPGPPGPPGPPG 1190

RESULT 2
S21626
collagen alpha 1(I) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626
R;Li, S.W.; Khillan, J.; Prockop, D.J.

Matrix Biol. 14, 593-595, 1994
A;Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I
A;Reference number: S57243
A;Accession: S57243
A;Molecule type: mRNA
A;Residues: 1-1453 <LIS>
A;Cross-references: UNIPROT:P11087; EMBL:U08020; NID:G470673; PIDN:AAA88912.1; PID:G470
R;Metaeranta, M.; Tonan, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A;Reference number: S16176; MUID:91274355; PMID:2054384
A;Accession: S16374
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1442-1453 <MET>
A;Cross-references: EMBL:X57981; NID:G50484; PIDN:CAA41046.1; PID:G50485
R;French, B.F.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A;Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A;Reference number: A23982; MUID:86137403; PMID:3841523
A;Accession: A23982
A;Molecule type: mRNA
A;Residues: 518-1128 <PRE>
A;Cross-references: GB:M14423; NID:G192261; PIDN:AAA37333.1; PID:G192262
R;Monson, J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A;Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for
A;Reference number: I49559; MUID:83141374; PMID:6298597
A;Accession: I49559
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 735-1130 <RES>
A;Cross-references: GB:M17491; NID:G192263; PIDN:AAA37334.1; PID:G192264
R;Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A;Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads
A;Reference number: I49557; MUID:84170331; PMID:6324198
A;Accession: I49557
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-25 <RE2>
A;Cross-references: GB:K01688; NID:G192246; PIDN:AAA37330.1; PID:G553881
R;Fenton, S.P.; Lanada, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A;Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A;Reference number: S39789; MUID:94092741; PMID:8268229
A;Accession: S39789
A;Molecule type: DNA
A;Residues: 180, 'E', 82-105, 'D', 107-185; 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-
R;Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
Mol. Cell. Biol. 14, 5950-5960, 1994
A;Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indire
A;Reference number: I48300; MUID:94344105; PMID:8065328
A;Accession: I48300
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 180, 'E', 82-105, 'D', 107-147 <REF>
A;Cross-references: EMBL:X54876; NID:G50486; PIDN:CAA38657.1; PID:G50487
C;Genetics:
A;Gene: COL1A1
A;Introns: 770/3; 788/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil, extracellular matrix; glycoprotein; heterotrimer; triple helix
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-151/Domain: amino-terminal propeptide #status predicted <PRO>
F;30-89/Domain: von Willebrand factor type C repeat homology <VWC>
F;152-1453/Product: collagen alpha 1(I) chain #status predicted <VAT>
F;1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 63.2%; Score 2205.5; DB 2; Length 1453;
Best Local Similarity 58.9%; Pred. No. 6e-104;
Matches 432; Conservative 22; Mismatches 162; Indels 117; Gaps 11;

QY 1 GPP-----GEPGPTGLPGPPGERGGPSRGGPPGADGVAGKPGPAGERGSGPPA 48
 |||||
 Db 447 GPPGPAGEGKRGARGEPGSLGPPGPPGERGGPSRGGPPGADGVAGKPGPAGERGSGPPA 506
 QY 49 GPKGSPGAGPGEAGLPGAKGLTGSPPSPDPDKTTPPPAGDGGPPGPPGARGQA 108
 |||||
 Db 507 GPKGSPGAGPGEAGLPGAKGLTGSPPSPDPDKTTPPPAGDGGPPGPPGARGQA 566
 QY 109 GVMGPPGKGAAGBPCKAGRGVPPGCAVGPAGKDGAGAGQAGPPGAGPAGERGEGCPA 168
 |||||
 Db 567 GVMGPPGKGTAGEPGKAGRGVPPGCAVGPAGKDGAGAGQAGPPGAGPAGERGEGCPA 626
 QY 169 GSPGFQGLPGPAGPGEAGKPGEGVPGDGLGAPGSPGAGE-----PGP----- 212
 |||||
 Db 627 GSPGFQGLPGPAGPGEAGKPGEGVPGDGLGAPGSPGAGE-----PGP----- 686
 QY 213 -----TGLPGLPGERGGPSRGGPPGADGVAGKPGPAGERGSGPPGARGSP 258
 |||||
 Db 687 GNNGAPGNDGAKGDTGAFAPGSPGAGPGLQMPGERSGAAGLPGPKGDRGDAGPKGADGSP 746
 QY 259 GE-----AGRPGEAGLPGAKGLTGSPPSPDP-----GKTGPPGAGDGRP 300
 |||||
 Db 747 CKDGARGITGPIGPPGAGAPGDKGAGSPGPPGPTGARGAPGDRGAGPPGAGPAGPP 806
 QY 301 GPPGPPGARGQ-----AGVMGPPGKGAAGEPKKAGRGVPPGPPGA----- 341
 |||||
 Db 807 GADQPGAKGPPGTGVKGDAGPPGAPGAPGPPGPIGNVAPGPKGPRGAAGPPGATGFP 866
 QY 342 -----VGPAGKDGAGAGQPPGAPG-----AGERGEGPAGSPGQGLPGPAGPAGGAGKP 393
 |||||
 Db 867 GAACKGVPPGSPGNAGPPGPPGVGKGGKPRGCTGAGPAGPGEVGPFGPPGAGEKGP 926
 QY 394 GEQGVPGDLGAPGSPGAGPPTGLPGLPGERGGPSRGGPPGADGVAGKPGPAGERGSP 453
 |||||
 Db 927 GADGAPGSPGPPGQGTAGQGVVGLPQGRGERGFPGLPSPGEPKQKSGSGSGRGP 986
 QY 454 GPAGP-----KSPGAGPGEAGLPGAKGLTGSPPSPDPDKTTPPPAGDGGPPGPPG 510
 |||||
 Db 987 GPMGPPGLAGPPGSGREGSPGAGSGPGRDGAAGKAGDGTGAPGPPGAPGAPGPPV 1046
 QY 511 GARQAGVMGPPGPKGAAGBPCKAGRGVPPG-----PGAV 546
 1047 GPAGKNGDRGTGAPGAPGPIGAGARGPAGPQGRGDKGTGEGDRTGKHRGFSGLQ 1106
 QY 547 GPAGKDGAGAGQPPGAPGAGPAGERGEGPAGSP-----GFQGLPGPAGPGEAGKPGPQGV 603
 |||||
 Db 1107 GPFSPSPGEPGQPSGASGAPGPPGSPGAGSGPCKDGLNGLPGIPGPPRGRGTGDSGPA 1166
 QY 604 GDLGAPGSPGAP 616
 |||||
 Db 1167 GPPGPPGPPGPPG 1179

RESULT 3

CGCHIS

collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)

C;Species: Gallus gallus (chicken)
 C;Date: 12-Aug-1991 #sequence revision 06-Jul-1982 #text_change 31-Mar-2000
 C;Accession: A90458; A90181; A02857
 R;Higberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyver, J.M.; Kang, A.H.; Gross, J.
 Biochemistry 21, 2048-2055, 1982
 A;Title: Amino acid sequence of chick skin collagen alpha1(I)-C88 and the complete prima
 A;Reference number: A90458; MUID:82231995; PMID:7093229
 A;Accession: A90458
 A;Molecule type: protein
 A;Residues: 1-1036 <HIG>
 A;Experimental source: skin
 A;Note: this is the latest in a series of papers from these workers elucidating the sequ
 R;Eyre, D.R.; Glimcher, M.J.
 Biochem. Biophys. Res. Commun. 48, 720-726, 1972
 A;Title: Evidence for a previously undetected sequence at the carboxyterminus of the alp
 A;Reference number: A90181; MUID:72243016; PMID:5047697
 A;Accession: A90181

A;Molecule type: protein
 A;Residues: 1037-1042 <EVR>
 A;Experimental source: skin

A;Note: residues 1037-1042 above correspond to the carboxyl end of the protein
 C;Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some
 C;Comment: Most of the prolines at the third position of the tripeptide repeating unit
 C;Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in po
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyrrolidone carboxylic acid; trimer;
 F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 61.7%; Score 2151.5; DB 1; Length 1042;
 Best Local Similarity 57.8%; Pred. No. 2.4e-101;
 Matches 424; Conservative 17; Mismatches 175; Indels 117; Gaps 10;

QY 1 GPP-----GEPGPTGLPGPPGERGGPSRGGPPGADGVAGKPGPAGERGSGPPA 48
 |||||
 Db 296 GPPGPAGEGKRGARGEPGSLGPPGPPGERGGPSRGGPPGADGVAGKPGPAGERGSGPPA 355
 QY 49 GPKGSPGAGPGEAGLPGAKGLTGSPPSPDPDKTTPPPAGDGGPPGPPGARGQA 108
 |||||
 Db 356 GPKGSPGAGPGEAGLPGAKGLTGSPPSPDPDKTTPPPAGDGGPPGPPGARGQA 415
 QY 109 GVMGPPGKGAAGBPCKAGRGVPPGCAVGPAGKDGAGAGQAGPPGAGPAGERGEGCPA 168
 |||||
 Db 416 GVMGPPGKGTAGEPGKAGRGVPPGCAVGPAGKDGAGAGQAGPPGAGPAGERGEGCPA 475
 QY 169 GSPGFQGLPGPAGPGEAGKPGEGVPGDGLGAPGSPGAGE-----PGP----- 212
 |||||
 Db 476 GADGAPGSPGPPGQGTAGQGVVGLPQGRGERGFPGLPSPGEPKQKSGSGSGRGP 535
 QY 213 -----TGLPGLPGERGGPSRGGPPGADGVAGKPGPAGERGSGPP----- 251
 |||||
 Db 536 GANAPGNDGAKGDTGAFAPGSPGAGPGLQMPGERSGAAGLPGPKGDRGDAGPKGADGSP 595
 QY 252 -----AGPKGSPGAGRP-----GEAGLPGAKGLTGSPPSPDPDKTTPPPGPPA----- 294
 |||||
 Db 596 GKDLGRLGTGPIGPPGAGAPGDKGAGSPGPPGPTGARGAPGDRGEPGPPGAPGAPG 655
 QY 295 -----GQDRFPGPPGPPGARGQAAGVMGPPGPKGAAGBPCKAGRGVPPG 336
 656 GADQPGAKGDTGADGAGKDGAGPPGAPGTCAPAGZVGVGAPGPKGARGSGAGPAGTGP 715
 QY 337 GPPGAVGAPGKDGAGAGQPPGAPGAGE-----RGEQGPAGSPGQGLPGPAGPGEAGKP 393
 716 GAAERVPGPPSGNIGLPGPPGPAKKGSKGPRGTGAPRBPBPAGPAGPBPPEKGP 775
 QY 394 GEQGVPGDLGAPGSPGAPGPTGLPGLPGERGGPSRGGPPGADGVAGKPGPAGERGSP 453
 776 GADGPIGAPGTPGQGTAGQGVVGLPQGRGERGFPGLPSPGEPKQKSGSGSGRGP 835
 QY 454 GPAGP-----KSPGAGPGEAGLPGAKGLTGSPPSPDPDKTTPPPAGDGGPPGPP 510
 836 GPMGPPGLAGPPGSGREGAPGAGPGRDGAAGKAGDGTGAPGPPGAPGAPGPPV 895
 QY 511 GARQAGVMGPPGPKGAAGBPCKAGRGVPPGPPGAVGAPGAKDGE-----AGAQ 558
 896 GPAGKNGDRGTGAPGAPGPIGAGARGPAGPQGRGDKGTGEGDRTGKHRGFSGLQ 955
 QY 559 GPP-----GPAGPAGERGEGPAGSPGQGLPGPAGPGEAGKPGPQGV 603
 |||||
 Db 956 GPPGPPGAPGEPGSPGASGAPGPPGSPGAGSGPCKDGLNGLPGIPGPPRGRGTGEGVP 1015
 QY 604 GDLGAPGSPGAP 616
 |||||
 Db 1016 GPPGPPGPPGPPG 1028

RESULT 4

CGRTIS

collagen alpha 1(I) chain - rat (tentative sequence) (fragments)

C;Species: Rattus norvegicus (Norway rat)
 C;Date: 13-Jul-1981 #sequence revision 13-Jul-1981 #text_change 31-Mar-2000

C;Accession: A90559; A90552; A92029; A90353; A90566; A90357; A90362; A90379; A91209; A91
R;Bornstein, P.
Biochemistry 8, 63-71, 1969
A;Title: Comparative sequence studies of rat skin and tendon collagen. II. The absence of
A;Reference number: A90559; MUID:69155173; PMID:5777344
A;Contents: CNBr0 and CNBr1
A;Accession: A90559
A;Molecule type: protein
A;Residues: 1-19 <BO1>
A;Experimental source: tendon
A;Note: sequences from skin and tendon appear to be identical
A;Note: the amino-terminal tetrapeptide may be removed by limited proteolysis during ext
R;Kang, A.H.; Bornstein, P.; Piez, K.A.
Biochemistry 6, 788-795, 1967
A;Title: The amino acid sequence of peptides from the cross-linking region of rat skin c
A;Reference number: A90552; MUID:67162268; PMID:5337886
A;Contents: CNBr1
A;Accession: A90552
A;Molecule type: protein
A;Residues: 5-19 <KAN>
A;Experimental source: skin
R;Bornstein, P.
J. Biol. Chem. 242, 2572-2574, 1967
A;Title: The incomplete hydroxylation of individual prolyl residues in collagen.
A;Reference number: A92029; MUID:67165368; PMID:4290711
A;Contents: CNBr2
A;Accession: A92029
A;Molecule type: protein
A;Residues: 20-55 <BO2>
A;Experimental source: skin and tendon
R;Butler, W.T.; Ponds, S.L.
Biochemistry 10, 2076-2081, 1971
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino a
A;Reference number: A90353; MUID:71263178; PMID:4327399
A;Contents: CNBr4
A;Accession: A90353
A;Molecule type: protein
A;Residues: 56-102 <BU1>
A;Experimental source: skin
R;Butler, W.T.
Biochemistry 9, 44-50, 1970
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. The co
A;Reference number: A90566; MUID:70085124; PMID:5411206
A;Contents: CNBr5
A;Accession: A90566
A;Molecule type: protein
A;Residues: 103-139 <BU2>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Bornstein, P.
Biochemistry 10, 4470-4478, 1971
A;Title: Structure of rat skin collagen alphas-1, 1971
A;Reference number: A90357; MUID:72136131; PMID:4335087
A;Contents: CNBr8
A;Accession: A90357
A;Molecule type: protein
A;Residues: 140-238 <BA1>
A;Experimental source: skin
R;Balian, G.; Click, E.M.; Hermodson, M.A.; Bornstein, P.
Biochemistry 11, 3798-3806, 1972
A;Title: Structure of rat skin collagen alphas-1, 1972
A;Reference number: A90362; MUID:73006942; PMID:4342027
A;Contents: CNBr8
A;Accession: A90362
A;Molecule type: protein
A;Residues: 239-418 <BA2>
A;Experimental source: skin
R;Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.
Biochemistry 13, 2946-2953, 1974
A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino a
A;Reference number: A90379; MUID:74271984; PMID:4366532
A;Contents: CNBr3
A;Accession: A90379
A;Molecule type: protein

A;Residues: 419-567 <BU3>
A;Experimental source: skin
R;Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.
Eur. J. Biochem. 37, 287-294, 1973
A;Title: Structural and immunogenic properties of a major antigenic determinant in neut
A;Reference number: A91209; MUID:74011954; PMID:4126850
A;Contents: CNBr6
A;Accession: A91209
A;Molecule type: protein
A;Residues: 568-651 <ST1>
A;Experimental source: skin
A;Note: this region probably corresponds to positions 949-1032 of the alpha 1(I) chain
A;Note: the major antigenic determinant (of neutral salt-extracted rat skin collagen) i
R;Stoltz, M.; Timpl, R.; Kuehn, K.
FEBS Lett. 26, 51-55, 1972
A;Title: Non-helical regions in rat collagen alphas-1.
A;Reference number: A91385; MUID:73049495; PMID:4636751
A;Contents: CNBr6
A;Accession: A91385
A;Molecule type: protein
A;Residues: 651-671 <ST2>
A;Experimental source: skin
A;Experimental source: skin
A;Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the sequ
A;Note: this region (residues 651-671 above) probably corresponds to positions 1032-105
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Comment: The order of the nine CNBr peptides in the alpha 1(I) chain of rat skin coll
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxy
F;1/Modified site: blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #stat
F;9/Modified site: allysine (Lys) #status experimental
F;103,424,547/Binding site: carboxylate (Lys) (covalent) #status experimental
F;103/Modified site: 5-hydroxylysine (Lys) #status experimental
F;424,547/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
Query Match 57.7%; Score 2011.5; DB 1; Length 671;
Best Local Similarity 59.5%; Pred. No. 1.8e-94;
Matches 396; Conservative 28; Mismatches 119; Indels 123; Gaps 14;
QY 1 GPPGPGTGLPGERGGGS-----RGPFGA-----DGVAGPKPAGERGSPGA 48
Db 29 GPGFAPGPGFQGGPGEPEGPGAGPMGPRGPPGPPGKNDGDEAGKGRGPRGPPG 88
QY 49 GPKGPGGEAGRPGE-----AGLPGAKGLT-----GSPGSPGPD---GKTGPPGPGAGD 93
Db 89 GARGLPCTAGLPGMKHRCFSLDCAKNTGPAKPGKPGSPGNGAGPQMGPRGLPGER 148
QY 94 GRPGGPPGARGQAGVMGFPKGAAGE-----PGKAGERGVPPGPAVAGPKDGEA 147
Db 149 GRGPPGSGAGRDDCAVGAAGFPPTGTPGPPGFGAAGKAGEAGPQARGSEGPQVR 208
QY 148 GAQGGPPGAPAGERGEQGP-----AGSPFGQLPFPAGPPEAGKPGGQ 192
Db 209 GEFPPGPAAGAPAGNPGADQCPGAKGANGAPGTAGAPGFFCARGSPGQPSGAPGPK 268
QY 193 GVPGLDGP-----GPSGA-----GFPCTGLPFPGERGGP 225
Db 269 GNSGEPAGPNKGDITGAKGPGPAGVQGGPPGPAEGEGRKARGEPGSLGPPGERGGF 328
QY 226 GRRGPGADGVAGPKPAGERSPGAPKSGSPGAGRPGEGAGLPGAKGLTSGSPSGPD 285
Db 329 GRRGPGADGVAGPKPAGERSPGAPKSGSPGAGRPGEGAGLPGAKGLTSGSPSGPD 368
QY 286 GKTGPPGAPQDGRPPGPPGARGQAGVMGFPKGAAGE-----GKAGERGVPPGPAVAGPA 345
Db 389 GKTGPPGAPGZBGRPGAPGPPGARGQAGVMGFPKGTAGEFGKAGERVPPGPAVAGPA 448
QY 346 GKDGAGAGGPPGAPGAPAGEREQGPAGSPGQGLPGPAGPPGEGAGKPGEQVPGDLGAP 405
Db 449 GKDGAGAGGAPGAPGAPAGEREQGPAGSPGQGLPGPAGPPGEGAGKPGZGVPGLGAP 508
QY 406 GPSGPAGEPGTGLPGERGGSPGSRGPPGADGVAGPKPAGERGSPGAPKSGPGEA 465

```

Db      509 GPSGARGE-----RGPFGERVQGPAGPRGN-----NGAPGBB 544
QY      466 GRPGEAGLFGKGLTSGSPGPGDKTGPFGPAQDGRPGPPGPGARGAGVGMFGPGPK 525
Db      545 GAKGDTGAPGAPGSGGAPGLZGMSGLZGPPGPGS---PGZGSPGASGAPGR---GPP 598
QY      526 GAAGFPGKAGRGVPGPCAVGAPGAGKAGCAQAGPAGPAGRGQAGPAGSGFPGGLP 585
Db      599 GSASPGKGBLGLPGIPGPRGRGTGBAGSPGPPGPGP-----P 640
QY      586 GPAGPP 591
Db      641 GPPGPP 646

RESULT 5
745467
collagen alpha 1(II) chain precursor [imported] - horse
N:Alternate names: type II collagen
C:Species: Equus caballus (domestic horse)
C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C:Accession: T45467
R:Richardson, D.W.; Dodge, G.R.
submitted to the EMBL Data Library, June 1996
A:Description: Cloning of equine type II collagen and modulation of its expression in eq
A:Reference number: 222977
A:Accession: T45467
A>Status: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: mRNA
A:Residues: 1-1418 <RIC>
A:Cross-references: UNIPROT:Q28396; EMBL:U62528; PIDN:AA805773.1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 55.7%; Score 1944; DB 2; Length 1418;
Best Local Similarity 52.1%; Pred. No. 7.3e-91;
Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY      1 GPPCEPPTGLPGPPGRRGGSGRPGGADGVAGPKGPGAGERSGPGAPGKSGPGAGRP 60
Db      156 GNPEGEPGPGVGMGPRGPPGPKGDDGEAGKPKSGERGGPPGQAGRGPPGPGLP 215
QY      61 -----GEAGLFGKGLTSGP---GSRGP-----DKTGTGPPGPA--- 90
Db      216 GVKCHRGVPLDGAKEGAGAPGVKGESGSPGNGSPGMPGRLPGERGRTGPAAGAR 275
QY      91 GQDRPPTGPPGARGAGVGMGPRGPKGAAGEPKAGER-----GVPGP- 134
Db      276 GNDGQPGPAGPPGPGVAGGPPGAPGAPGAGKAGGAGPTGARGPEGAQGPGRGPPGPGPA 335
QY      135 -----PGAVGPAKGDGAGAGQPPGAPGAGEGQGP-----AGSPGQGLP 177
Db      336 GAAGNCTDGIKAGKAGAGAPGIAGAPGFPGRGPPGQATGTLKQGTGEPGATGFK 395
QY      178 GPAGPPEAKPBEQGVFDLAPGSPGAPGEGPTGLPFPGERGGPSRFPFGADGVA 237
Db      396 GEQPKGPEPAGPQGAAPGAPGEGKRGARGEPGAGVPVPPGERGAPNRRGFPFGDGLA 455
QY      238 GPKGPAGERSGPGAPGKSGFAGRPGGAGLFGKGLTSGSPGPDGKTGPPGAGOD 297
Db      456 GPKGAPGERGSPGLAGPKGANGPGRGPEGLPFGARGLTGRPDAGQKVPSGAPGD 515
QY      298 GRGPPGPPGARGAGVGMGPRGPKGAAGEPKAGRGVPGPPGAPGAPKDGEGAGAQGPP 357
Db      516 GRFGPPGPPGARGQPGVGMGPRGPKGANGEPKAGRGKGLPCAPGLRGLPKDGTGTAAAGPP 575
QY      358 GPAGPAGERGEQGPAGSPGQGLPGGAPGPGGKAGKPGEGVPGDLGAPGSPGAGE---P 414
Db      576 GPAGPAGERGEQAGPAGSPGQGLPGGAPGPGGKAGKPGEGVPGGAGAPGLVPRGERGPP 635
QY      415 GPTGLP-----GPFGERGGPSRFPFGADGVAGPKGPA 447
Db      636 GERGSPGAOGLQARGLPPTGTDGPKGASGAPGAPGAPGQPPGGLQMPGERGAAGIAGPK 695

```

```

QY      448 GERGSPGAPGKSGPGE-----AGRPGEAGLFGKGLTSGSPGSPGD 489
Db      696 GDRGVDGKGPBAGKDGKGRGLTGPIGPPGAPGANGKGEVGGPPGAGTAGARGAPGER 755
QY      490 KGTGPPGAGQDGRGPPGPGAR-----COAGVMGPPGPKGAAGER-----GKA 534
Db      756 GTGTPGPPGAPGAPGADGQPCAKGEGQGEAGQKDGAGAPGQGGSGAPGPGPGTGVTPGK 815
QY      535 GERGVGPPGPA-----VGPAGKDGEGAGQAQPPGPA---GPAERGEQGPAGSPGFQ 582
Db      816 GARGAQPPGATGFPGAAGRVGPPGSGNGPPGPPGPGSGKDGKAGGDSGPPGGRAGDP 875
QY      583 GLFGPAGPPGAGKPGEGQGVFGDLGAPGSPGAP 616
Db      876 GLQGPAGPPGKGEFGDDGSPGDPGPPGQGLAG 909

RESULT 6
CGHU6C
collagen alpha 1(II) chain precursor [validated] - human
N:Alternate names: procollagen alpha 1(II) chain
N:Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen
C:Species: Homo sapiens (man)
C>Date: 28-May-1986 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C:Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S63
7450; I37251; I37252; I37253; I37254; I55338; I55335; I61910
R:Ryan, M.C.; Sieraski, M.; Sandell, L.J.
Genomics 8, 41-48, 1990
A:Title: The human type II procollagen gene: identification of an additional protein-cod
A:Reference number: A38513; MUID:91184811; PMID:2081599
A:Accession: A38513
A:Molecule type: DNA
A:Residues: 1-103 <RYA>
A:Cross-references: UNIPROT:P02458; UNIPROT:Q14042; UNIPROT:Q16672; UNIPROT:Q12985; UNIP
R:Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
Nucleic Acids Res. 17, 9473, 1989
A:Title: Nucleotide sequence of the full length cDNA encoding for human type II procolla
A:Reference number: S06715; MUID:90067946; PMID:2587267
A:Accession: S06715
A:Molecule type: mRNA
A:Residues: 1-28, 'R', '99-1487 <SU2>
A:Cross-references: EMBL:X16468; NID:g29515; PIDN:CAA34488.1; PID:g29516
A>Note: alternative splice form 1
R:Vikkula, M.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.
Biochem. J. 285, 287-294, 1992
A:Title: Structural analysis of the regulatory elements of the type-II procollagen gene.
A:Reference number: S24270; MUID:92344585; PMID:1637314
A:Accession: S24270
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-28 <VIK>
A:Cross-references: EMBL:X58709; GB:S40537; NID:g35659
A>Note: this translation is not annotated in GenBank entry HSFROCOEL, release 111.0
R:Jimenez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
Gene 44, 11-16, 1986
A:Title: Promoter region of the human pro-alpha-1-(II)-collagen gene.
A:Reference number: A24828; MUID:87031574; PMID:3021582
A:Accession: A24828
A:Molecule type: DNA
A:Residues: 1-8, 'T', '10-28 <NUN>
A:Cross-references: GB:M25698; NID:g180872; PIDN:AAA52051.1; PID:g553237
R:Balmain, C.T.; Reginaldo, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
Biochem. J. 262, 521-528, 1989
A:Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(II)
A:Reference number: S06496; MUID:90026318; PMID:2803268
A:Accession: S06496
A:Molecule type: mRNA
A:Residues: 7-28, 'R', '99-157, 'P', '159-440, 'G', '442-456, 'E', '458-640, 'A', '642-831, 'PA', '834, 'F'
A:Cross-references: EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID:g30041
A>Note: alternative splice form 1
R:Ryan, M.C.; Sandell, L.J.
J. Biol. Chem. 265, 10334-10339, 1990

```

A;Title: Differential expression of a cysteine-rich domain in the amino-terminal propept
A;Reference number: A35428; MUID:90285153; PMID:2355003
A;Accession: A35428
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 27-81, 'L', 83-103 <RYA2>
A;Note: alternative splice form 2; splicing appears to be under developmental regulation
R;Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, P.
Genomics 4, 438-441, 1989
A;Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf
A;Reference number: A30147; MUID:99233138; PMID:2714801
A;Accession: A30147
A;Molecule type: DNA
A;Residues: 104-157, 'P', 159-236 <SUM>
A;Cross-references: GB:J03065; GB:M23660; GB:M25656; GB:M25730; GB:M32168; GB
P;Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 6585-6568, 1990
A;Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pri
A;Reference number: A94227; MUID:90370826; PMID:1975693
A;Accession: A33116
A;Molecule type: DNA
A;Residues: 171-172, 'C', 174-175 <ALA>
A;Note: mutant sequence from a family with primary generalized osteoarthritis
R;Diab, M.; Wu, J.J.; Eyre, D.R.
Biochem. J. 314, 327-332, 1996
A;Title: Collagen type IX from human cartilage: a structural profile of intermolecular c
A;Reference number: S64673; MUID:96195147; PMID:8660302
A;Accession: S64674
A;Molecule type: protein
A;Residues: 188-189, 'X', 191-195, 1224-1230, 'X', 1232-1236 <DIA>
R;Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,
Eur. J. Biochem. 234, 125-131, 1995
A;Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil can
A;Reference number: S63514; MUID:96096730; PMID:8529631
A;Accession: S63514
A;Molecule type: protein
A;Residues: 243-251, 575-590, 756-763, 'X', 765-779 <FBA>
R;Tiller, G.E.; Weis, M.A.; Polunbo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,
Am. J. Hum. Genet. 56, 388-395, 1995
A;Title: An RNA-splicing mutation (G-to-SIVS20) in the type II collagen gene (COL2A1) in a
A;Reference number: I38867; MUID:95150028; PMID:7847372
A;Accession: I38867
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 440, 'G', 442-456, 'E', 458-480, 'P', 482-509 <TLLA>
A;Cross-references: EMBL:U15195; NID:9557053; PIDN:AA560370.1; PID:9557054
R;Ramirez, F.
submitted to the EMBL Data Library, December 1988
A;Reference number: S04892
A;Accession: S04892
A;Molecule type: mRNA
A;Residues: 501-676, 'A', 678-783, 'A', 785-831, 'PA', 834, 'P', 836-1214 <RAM>
A;Cross-references: EMBL:X13783; NID:930037; PIDN:CAA32030.1; PID:930050
R;Viikkula, M.; Peltonen, L.
FEBS Lett. 250, 171-174, 1989
A;Title: Structural analyses of the polymorphic area in type II collagen gene.
A;Reference number: S05000; MUID:89325561; PMID:2753125
A;Accession: S05000
A;Molecule type: DNA
A;Residues: 630-640, 'A', 642-785 <VIK2>
A;Cross-references: EMBL:X16158; NID:929951; PIDN:CAA34278.1; PID:91335018; PIDN:CAA3427
PIDN:CAA34283.1; PID:91335023; PIDN:CAA34284.1; PID:91335024
R;Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D
J.; Biol. Chem. 267, 22522-22526, 1992
A;Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain pro
A;Reference number: A44309; MUID:93054548; PMID:1429602
A;Accession: A44309
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA; mRNA
A;Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'O', 1038-1052, 'E', 1054-1068, 'T',
A;Cross-references: GB:L00977; NID:9180812; PIDN:AB23914.1; PID:9358774
A;Note: sequence extracted from NCBI backbone (NCBI:117273); parts of this sequence we
A;Note: this translation is not annotated and this publication is not cited in GenBank &

A;Note: mutant sequence associated with perinatal lethal hypochondrogenesis
R;Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A;Title: Random duplication within a type II collagen gene (COL2A1) exon in an individu
A;Reference number: S16502; MUID:90251662; PMID:2339128
A;Accession: S16502
A;Molecule type: DNA
A;Residues: 1164-1184, 'GPSKQDANGIPGP', 1185-1199 <TIL2>
A;Cross-references: EMBL:M37126; NID:9180808; PIDN:AAA52037.1; PID:9180809
A;Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
R;Chesah, K.S.B.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
A;Title: Identification and characterization of the human type II collagen gene (COL2A1
A;Reference number: A02858; MUID:85190534; PMID:3857598
A;Accession: A02858
A;Molecule type: DNA
A;Residues: 1032-1056, 'N', 1058-1069, 'T', 1070-1487 <CHE>
A;Cross-references: GB:J00116; NID:9180395; PIDN:AAA51997.1; PID:9180396
R;Elima, K.; Vuorio, T.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A;Title: Determination of the single polyadenylation site of the human pro-alpha-1(II)
A;Reference number: A27280; MUID:88067771; PMID:2825137
A;Accession: A27280
A;Molecule type: DNA; mRNA
A;Residues: 1175-1487 <ELI>
A;Cross-references: EMBL:X06268; NID:930096; PIDN:CAA29604.1; PID:930097
A;Experimental source: fetal epiphyseal cartilage
R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 237, 923-925, 1986
A;Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A;Reference number: A57033; MUID:87099927; PMID:3800925
A;Accession: A57033
A;Molecule type: protein
A;Residues: 'XE', 1244-1246, 'N', 1248, 'X', 1250-1265, 1295-1305, 1395-1408 <VAN>
A;Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal pro
R;Strom, C.M.; Upholt, W.B.
Nucleic Acids Res. 12, 1025-1038, 1984
A;Title: Isolation and characterization of genomic clones corresponding to the human ty
A;Reference number: A21733; MUID:84118798; PMID:6320112
A;Accession: A21733
A;Molecule type: DNA
A;Residues: 1245-1295 <STR1>
A;Cross-references: EMBL:X00339; EMBL:X00298; NID:9394699; PIDN:CAA25092.1; PID:9437897
A;Accession: B21733
A;Molecule type: DNA
A;Residues: 894-909, 'PE' <STR2>
A;Cross-references: GB:K01785; NID:930035; PIDN:CAA5082.1; PID:91335032
R;Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A;Title: Isolation and partial characterization of genomic clones coding for a human pr
gene
A;Reference number: A24561; MUID:86104139; PMID:3002437
A;Accession: A24561
A;Molecule type: DNA
A;Residues: 1296-1358 <NUN2>
A;Cross-references: GB:M12048; NID:9180017
A;Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
A;Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with th
R;Sangiorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez
Nucleic Acids Res. 13, 2207-2225, 1985
A;Title: Isolation and partial characterization of the entire human pro alpha 1(II) col
A;Reference number: I37249; MUID:85215609; PMID:2987845
A;Accession: S59491
A;Molecule type: DNA
A;Residues: 7-28, 'R', 99-114, 541-578, 786-802, 1055-1056, 'N', 1058-1068, 'T', 1070-1109, 1200-
A;Accession: I84453
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 7-28 <GAN2>
A;Cross-references: GB:M23759; NID:9180845; EMBL:X03320; GB:M24938; NID:930104
A;Note: the GenBank PID is based on an incorrect reading frame
A;Accession: I37250
A;Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 541-560 <SANK>
A:Cross-references: EMBL:X02378; GB:M23870; NID:G30107; PIDN:CAA26227.1; PID:G929621
A:Accession: 137251

Query Match 55.7%; Score 1943; DB 1; Length 1487;
Best Local Similarity 52.1%; Pred. No. 8.5e-91;
Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

```
QY 1 GPPCEPPTGLPGPPGSGSGFFGADGVAGPKPAGERSPPGAPGKSGPSGAGRP 60
DB 225 GNPCEPGEVSGPMGPRGPPGPKGDDGEAGPKGKAGERFPFGQARGFPPTGFLP 284
QY 61 -----GEAGLPGAKGLTGSP---GSPGP-----DGKTGPPGPA--- 90
DB 285 GVKGHRGYPGLDGAKEAGAGVKGESGSGENGSPGMPGRGLPGERGTGAGAGAR 344
QY 91 QDGRPPGPPGARGAQGVWGF-----GPKGAAGEPKKAGERGVPGPAGVGA 141
DB 345 GNDQGPAGPPGPPGAGGPPGAPCAKGEAGTGAEGAGQPRGPPGTPGSPGA 404
QY 142 -----GKDGEAGAGPPGAPGAGEGOGPAGSPGFGQLPGPAGPPGAGKPGGQVP 195
DB 405 GASNPDTGLPGKAGSAGAPGATGAPFPGRPPDPQATGLPKGTGAPGATGPK 464
QY 196 GD---LGAPGSPGAPGPTGLP-----GPPGERGPGSRGPPGADGVA 237
DB 465 GEQGPKEGPPGAPGQAGPPGAGEGKAGRGPPGVPIGPPGERGAPGNRGPFGDGLA 524
QY 238 GPKPAGERSPPGAPGKSGPGEAGRGEGAGLPGKAGLTGSPGSPGDPGKTGPPGAG 297
DB 525 GPKGAPGERGSPGLAGPKGANGDPGRGEPGLPGARGLTRPGRDAGPGKVGSPGAPGD 584
QY 298 GRPPGPPGARGAQGVWGFPGKGAAGEPKKAGERGVPGPAGVGPAGKDGAGAGQGP 357
DB 585 GRPPGPPGARGAQGVWGFPGKGAAGEPKKAGERGVPGPAGVGPAGKDGAGAGQGP 644
QY 358 GPAGPAGERGEGPAGSPGQGLPGPAGPPGEGAGKPGEGQVPGDLAGPSPGAGE--- 413
DB 645 GPAGPAGERGEGPAGSPGQGLPGPAGPPGEGAGKPGEGQVPGDLAGPSPGAGE--- 704
QY 414 -----PGTGLP-----GPPGERGPGSRGPPGADGVAAGKGA 447
DB 705 GERSPGAQGLQGRGLPGTGTGDKGASGAPGAPGAPGAPGLQMPGERGAGTAGPK 764
QY 448 GERGS-----PGPAGPKSGPGEAGRPGEAGLPGAKGLTGSP 483
DB 765 GDRGVDGEGKEGAPGKDGGRGLTGTPTGPPGAPAGANGKEGVPGPAGSAGARGAPGR 824
QY 484 GSPPGDKTGPAGQDGRGPPGPPGAGGQAGVWGFPGPKGAAGEP-----GKA 534
DB 825 GETGPTGTGIAGPPGADGQGAKEGEGEAGQDAGAPGQPSGAPGQGTGVTGPK 884
QY 535 GERGVPPGPA-----VGPAGKDGAGAGQPPGPA---GPAGERGEGPAGSPGFG 582
DB 885 GARGAQPPGATGPPGAAGRVGPPGSGNPNPDPGPPGSKDGPKGARGDSPPGRAGER 944
QY 583 GLCPGAPGPEAGKPGEQVPGDLGAPGSPGAP 616
DB 945 GLQGPAGPPEKEGPPGDDGSGAEGPPGPOGLAG 978
```

RESULT 7
A41182
collagen alpha 1(II) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C:Accession: A41182; A44885
R:Metasera, M.; Tomar, D.; de Crombrughe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A:Reference number: A41182; MUID:91358489; PMID:1865613
A:Accession: A41182

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA
A:Residues: 1-1419 <MET>
A:Cross-references: GB:M65161
R:Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
Development 111, 945-953, 1991
A:Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilag
A:Reference number: A44885; MUID:91347939; PMID:1879363
A:Accession: A44885
A:Molecule type: DNA
A:Residues: 1-28 <CHE>
A:Cross-references: GB:S63190; NID:G234368; PIDN:AA819627.1; PID:G234369
A:Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBI:63192)
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime
F1191-1419/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 55.4%; Score 1932; DB 2; Length 1419;
Best Local Similarity 51.9%; Pred. No. 2.9e-90;
Matches 391; Conservative 36; Mismatches 189; Indels 138; Gaps 14;

```
QY 1 GPPCEPPTGLPGPPGSGSGFFGADGVAGPKPAGERSPPGAPGKSGPSGAGRP 60
DB 157 GNPCEPGEVSGPMGPRGPPGPKGDDGEAGPKGKAGERGLPFGQARGFPPTGFLP 216
QY 61 -----GEAGLPGAKGLTGSP---GSPGP-----DGKTGPPGPA--- 90
DB 217 GVKGHRGYPGLDGAKEAGAGVKGESGSGENGSPGMPGRGLPGERGTGAGAGAR 276
QY 91 QDGRPPGPPGARGAQGVWGF-----GPKGAAGEPKKAGERGVPGPAGVGA 141
DB 277 GNDQGPAGPPGPPGAGGPPGAPCAKGEAGTGAEGAGQPRGPPGTPGSPGA 336
QY 142 GKDE-----AGAQQPPGAPGAGEGEGQP-----AGSPGFGGLP 177
DB 337 GASNPCTGTGPKAGSAGAPGAGPFGPPGPRGPPGQATGFLGPKGQAGPPIAGPK 396
QY 178 GPAGPAGERGEGPAGSPGQGLPGPAGPPGEGAGKPGEGQVPGDLAGPSPGAGE--- 237
DB 397 GDQGPKEGTPGAPGQAGPPGAGEGKAGRGPPGAGPIGPPGERGAPGNRGPFGDGLA 456
QY 238 GPKPAGERSPPGAPGKSGPGEAGRPGEAGLPGKAGLTGSPGSPGDPGKTGPPGAGQD 297
DB 457 GPKAPGERGSPGLAGPKGANGDPGRGEPGLPGARGLTRPGRDAGPGKVGSPGAPGD 516
QY 298 GRPPGPPGARGAQGVWGFPGKGAAGEPKKAGERGVPGPAGVGPAGKDGAGAGQGP 357
DB 517 GRPPGPPGARGAQGVWGFPGKGAAGEPKKAGERGVPGPAGVGPAGKDGAGAGQGP 576
QY 358 GPAGPAGERGEGPAGSPGQGLPGPAGPPGEGAGKPGEGQVPGDLAGPSPGAGE--- 413
DB 577 GPSGAGERGEGPAGSPGQGLPGPAGPPGEGAGKPGEGQVPGDLAGPSPGAGE--- 636
QY 414 -----PGTGLP-----GPPGERGPGSRGPPGADGVAAGKGA 447
DB 637 GERSPGAQGLQGRGLPGTGTGDKGASGAPGAPGAPGAPGLQMPGERGAGTAGPK 696
QY 448 GERSPGAPGKSGPGE-----VGPAGKDGAGAGQPPGPA---GPAGERGEGPAGSPGFG 489
DB 697 GDRGVDGEGKEGAPGKDGGRGLTGTPTGPPGAPAGANGKEGVPGPAGSAGARGAPGR 756
QY 490 KTGTPPAGQDGRPPGPPGARG-----GQAGVWGFPGPKGAAGEP-----GKA 534
DB 757 GETGPPGAPGAPGADGQGAKEGEGEAGQDAGAPGQPSGAPGQGTGVTGPK 816
QY 535 GERGVPPGPA-----VGPAGKDGAGAGQPPGPA---GPAGERGEGPAGSPGFG 582
DB 817 GARGAQPPGATGPPGAAGRVGPPGSGNPNPDPGPPGSKDGPKGARGDSPPGRAGDP 876
QY 583 GLCPGAPGPEAGKPGEQVPGDLGAPGSPGAP 616
DB 877 GLQGPAGPPEKEGPPGDDGSGAEGPPGPOGLAG 910
```


C;Comment: The order of the eight CNBr peptides in the alpha 1(I) chain of bovine skin d
9, 149, 268, and 217 residues.

C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer;
F;1/Modified site: pyroglutamic acid (Gln) #status experimental

Query Match 54.4%; Score 1898; DB 1; Length 779;
Best Local Similarity 53.0%; Pred. No. 9.6e-89;
Matches 384; Conservative 22; Mismatches 210; Indels 108; Gaps 10;

```
QY 1 GPPGPGTGLPDPGGRGGPS-----RGFPADGVAGPKGAGRGSGSPGAGKGP 54
  |||||
DB 29 GPPGAPGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 88
  |||||
QY 55 GEAGRPGEAGLPGAKGLTSGSPGPGKTPGPGAGQDGRPGPPGPGARGQAGVMGPP 114
  |||||
DB 89 GARGLPTAGLPGWKHGRGSLDCAKGDAGPAGPKGPGSPGNGAPGQNGRGLPGFP 148
  |||||
QY 115 GPKAAGEPGKAGRGVPGPPGAYGAGKDGAGAGAGQPPGPGAGPAGRGQGPAGSPGFQ 174
  |||||
DB 149 GPKAAGEPGKAGRGVPGPPGAYGAGKDGAGAGAGQPPGPGAGPAGRGQGPAGSPGFQ 208
  |||||
QY 175 GLPGPAGPGEAGVGEQGVGPDGLGARGSPGAGE-----PGP----- 212
  |||||
DB 209 GLPGPAGPGEAGVGEQGVGPDGLGARGSPGAGE-----PGP----- 268
  |||||
QY 213 -----TGLPGPPGGRGPGSGRFPAGDGVAGPKGAGRGSGS----- 248
  |||||
DB 269 GNDGAKDAGAPAGPSGAGLQGMFGERGAAGLPGKCDRGDAGPKGADGAPGXGVR 328
  |||||
QY 249 -----PGPAGKPGSAGRGAGLPGCAKGLTSGSPGPDGKTGPPGAGQDGR- 299
  |||||
DB 329 GLTGP1GPPGAGAPGKAGSPGAGTGCAGDGRGEPGPPGAGPAGPPGADGQFGAK 388
  |||||
QY 300 -----PGPPGPGARGQAGVMGFPKGAAGEPGKAGRGVPGPPGAVGPA 345
  |||||
DB 389 GEPGDAGKDGAPPGPAGPAGPPGPIGVGAPGKAGSAGPPGATGPPGAGRVGPP 448
  |||||
QY 346 KDGEGAGQGPAGPAGE-----RGEGPAGSPGFGQLPGAPGPPGAGKPGGQVPGDL 402
  |||||
DB 449 GPSNAGPAGPPGPGAGKSGKPRGPGTGPAGRGPGGPPGPGAGKAPGADGAGAP 508
  |||||
QY 403 GAGPSPGAPGPGTGLPGRGPGSGRFPAGDGVAGPKGAGRGSPGAPG---K 459
  |||||
DB 509 GTPGPGIAGQGVVGLPGRGGRGFPGLPGSPGEPKQGPSASGERGPPGPMGPGLA 568
  |||||
QY 460 GSPGAGRPGEAGLPGAKGLTSGSPGPGKTPGPGAGQDGRPGPPPGARGQAGVM 519
  |||||
DB 569 GPPGESREGAPGAGSPGSDGSGAGKDRGETGAGAPGPPGAPGAPGVPAGKSGDR 628
  |||||
QY 520 GPPGPKAGEPKAGRGVPGPPGAVGAPKAGE-----AGAGPPG----- 563
  |||||
DB 629 GETGAPGPIGVGPGAGARGPAGPGPRGKGZTCZGBRGIKHGRGFSGLQGPSPGSP 688
  |||||
QY 564 -----AGPAGRGPGGAGSP-----GFGGLPGPAGPGEAGKPGGQGVGPDGLGAPGS 612
  |||||
DB 689 GEQGPSAGSPAGPRGPPGAGSPGKDLGLGPGIOPFRGTGADGAPGPPGPPGP 748
  |||||
QY 613 GPAG 616
  |||||
DB 749 GPPG 752
```

RESULT 10

B40333
collagen alpha 1(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: B40333
R;Su, M.W.; Suzuki, H.R.; Bleker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em

A;Reference number: A40333; MUID:92011898; PMID:1918153

A;Accession: B40333

A;Status: Preliminary

A;Molecule type: mRNA

A;Residues: 1-1486 <SUA>

A;Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M63595

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <VMC>
F;1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 54.2%; Score 1890.5; DB 1; Length 1486;
Best Local Similarity 54.6%; Pred. No. 3.6e-88;
Matches 374; Conservative 40; Mismatches 196; Indels 75; Gaps 10;

```
QY 1 GPPGPGTGLPDPGGRGGPSRFPAGDGVAGPKGPA-----GERGSGPAGPK 51
  |||||
DB 269 GPPGPGGARGPFGTGLPGVGHRYGLDCAKGAAGAKGEGGATGAGSGPMPGR 328
  |||||
QY 52 GSPGAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGQAGVM 111
  |||||
DB 329 GLPGRGGRGSGGAAGAR---GNDGLPGP---AGPPGVPAGAPGPPGAPSGXGAGPT 382
  |||||
QY 112 GPPGKGAAGPFGKAGRGVPGPPGAVGAGKDGAGAGQPPGPGAGRGGEQGPAGSP 171
  |||||
DB 383 GARGPAGAGRGSGSGTTPGSPGAGAGNPGTGTGPAKSGSGGPGTAGAPFPFPGRP 442
  |||||
QY 172 GFQGLPGPAGPPGAGPKP-----GEOQVPGDGLGAPGS---GPA-----GEFGPT 213
  |||||
DB 443 GPQATGTLPGKGTGDPGVAGFKGEGQKGEICSGAGPQAGPAGGEGKARGEGFAA 502
  |||||
QY 214 GLPFPGRGPGGSRGPPGADGVAGPKGAGRGSPGAGPKSGPAGRGPPGAGLPGAK 273
  |||||
DB 503 GPNPGRGAPGVRGPPGQDGLAGPKAGRGVPLGGPKGNGDPGPRPGLPGAR 562
  |||||
QY 274 GLTSPGSGPDGKTGTPPGPAGQDGRPGPPGARGQAGVMGFPKGAAGEPGKAGER 333
  |||||
DB 563 GLTGPADGAPGKVGSGSGEDGRFGPPGQARGQPGVMGFPGPKGANGEPKAGEK 622
  |||||
QY 334 GVPGPGAVGPKXDGAGAGAGPPGPGPAGPAGRGEGQGPAGSPGQGLPGPAGPPGAGPK 393
  |||||
DB 623 GLVAGPLGLPGKDGSTGSGQNGPAGPAGRGEGQGPAGSPGQGLPGPAGPPGSGGK 682
  |||||
QY 394 GEQGVGPDGLGARGSPGAGE---PGPTGLPGPGRGSGRSPGPPGADGVAGPKGAGER 450
  |||||
DB 683 GDQGVGAGAGPLVLRPRGERGFPGERGSGQGLGPRGLGPTGTDGPKGASGSPGPN 742
  |||||
QY 451 GSPGAPGKSGPAGRGAGEAGLPGCAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGPP 510
  |||||
DB 743 GAQGPGLQGMFGERGAAGISGPKGDRGTGKPEGASGKDGSRGLTGTPIGPPGAPGN 802
  |||||
QY 511 GARGQAGVMGFPKGAAGEPGKAGRGVPGPPGAVGAPKAGE-----GEA 555
  |||||
DB 803 GEKESGSGPPGIVGARGAPGDRGNGPPGPFAGPPGADGQSGLKGDQGESGQKGA 862
  |||||
QY 556 GAQGPMPGA-----GPAGRGEGQGPAGSPGFGGL-----PGPAGPP 591
  |||||
DB 863 GAPGQPSGAPPGQGTGTVFGPKGARGAGQAGPAGATGFPGAAGRVGTGPNNGNPGPP 922
  |||||
QY 592 GEAKPGEQGVGPDGLGAPGSPGAP 616
  |||||
DB 923 GSAGKEGPKVRGADGPPGAGDFG 947
```

RESULT 11

A40333
collagen alpha 1'(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
C;Accession: A40333
R;Su, M.W.; Suzuki, H.R.; Bleker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em

C:Reference number: A40333; MUID:92011890; PMID:1918153
A:Accession: A40333
A:Status: nucleic acid sequence not shown
A:Species: Mus musculus (house mouse)
A:Molecule type: mRNA
A:Residues: 1-1492 <SUA>
A:Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M3596
A:Note: this sequence is presented as substitutions relative to another sequence in a file
es they replace; the appropriate interpretation of the sequence figure was reconstructed
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:37-96/Domain: von Willebrand factor type C repeat homology <WVC>
F:1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 54.0%; Score 1883; DB 2; Length 1492;
Best Local Similarity 49.1%; Pred. No. 8.6e-88;
Matches 379; Conservative 46; Mismatches 185; Indels 162; Gaps 13;
QY 1 GPPGEPGTGLGPPGGERGGSGRFGCAD-----GVAGPKGAGGSGRFGPK 51
DB 272 GPPGQAGGPPGPPGPPGVKXGYYGLDGGKGEAGAAKAGSGAGAGAPGMPGR 331
QY 52 GSPGEARPGAGLPGAGKLTGSGSPGPDGKTGPPGAGQDGRPGPPGFGARGAGVM 111
DB 332 GLPGERGRPGSSGAAGAR---GNDGLRGP---AGPPGVPAGAPGPPGAPGSKGAGPT 385
QY 112 GPPGPKGAGBPGKAGRGVPGP-----PGAVGPKDGEAGAGQPPGAGPA 159
DB 386 GARGPEGAQGRGSGTPGSPGSGASGNPOTDIPGAKGSSGASGTAGAPGPPGRGPP 445
QY 160 GEREGQP-----AGSPFGGLPCPAGPCEAGKPCGQVPGDLGAPGSGPAGRGPT 213
DB 446 GQATGPLGPKGTGTPGAGTAGEHPKGEIGSAPQAPGAPGAGEGKRGAGPGAA 505
QY 214 GLPFPGERGPPGSRGPPGADGVAGPKGAPAGERGSPGAPKSGPSGAPGAGLPK 273
DB 506 GPLFPGERGAPGNRFPQDGLAGPKGAPGERGVPLGLGGPKGNGDPGRPGFPLPGAR 565
QY 274 GLTSPGSGPDGKTGTPPAGODGRPGPPGAPGAPGAPGAPGAPGAPGAPGAPGAP 333
DB 566 GLTGRPDAGQGVKSGSAGEDGRPGPPQARGQPVGMGPPGKANGEPGKAGX 625
QY 334 GVPFPAGVAGPKDGEAGAGQPPGAPGAGERGEGQPPGPGQGLPGPAGPGEAGKP 393
DB 626 GLGAPGLRGLPKDGTGTAQGPNGPAGPAGERGEGQPPGSPGQGLPGPPGSGEGKP 685
QY 394 GEQVPGDL-----GAPGSGPA 411
DB 686 GDQVPCBAGAGLVGPRGERGFTGERGSGPQGLGSRGLPPTGDPKAGATGPGSPN 745
QY 412 GEPGTGLPSPGPPGSG-----GP-----GSRGPPGADGVAGPKGPA 447
DB 746 CAQPPGLQGVGGERGAGISGPKDGRDTEKPEGAPKDGSRGLTGPLGPPGAPGN 805
QY 448 GERGSGPAGP-----KGSFGRAGRGEA 471
DB 806 CEKSGSPGPPGIVGARGAPGDRGNGPPGAPGAPGAPGAPGAPGAPGAPGAPGAP 865
QY 472 GLPKAGLTGSPGPPGPKTGP-----PGPAGODGRGPP---PGPPGAR 513
DB 866 GAPGQSGSGAPPGPGTGVNPGKARGAQGAPGATGPPGAGRGVTPGNGNPPGPPPP 925
QY 514 QGAVVMGPPGKGAAGEPKAGERGVPPGPAVAGPKDGEAGAGQPPGAPGAGERGEQ 573
DB 926 GSAGKEGKGVREGDAGTGRAGDPLQGPAGAPGKGEPEGPDGSPGDPGSPGQGLSGNR 985
QY 574 GPAGSP-----GQGLPCPAGPGEAGKP---GEQVPGDLGAPGSGPAG 616
DB 986 GIVGLPGQGRGPPGLGPPGPPGKGGSGDRGPPGPPGPPGPPGPPGPPGPPGPP 1037

RESULT 12
S59856
collagen alpha 1(III) chain precursor - mouse

C:Species: Mus musculus (house mouse)
C>Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text_change 09-Jul-2004
C:Accession: S59856; S62120; S16373
R:Toman, P.D.; de Crombrughe, B.
Gene 147, 161-168, 1994
A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA
A:Reference number: S59856; MUID:95011609; PMID:7926795
A:Accession: S59856
A:Molecule type: DNA
A:Residues: 1-1464 <TOM>
A:Cross-references: UNIPROT:P08121; EMBL:X52046
R:Toman, D.
submitted to the EMBL Data Library, November 1994
A:Reference number: S62120
A:Accession: S62120
A:Molecule type: DNA
A:Residues: 1-866, 'G', 868-1464 <TOA>
A:Cross-references: EMBL:X52046; NID:9575321; PIDN:CAA36279.1; PID:9575322
R:Metzger, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-249, 1991
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16373
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1464 <MET>
A:Cross-references: EMBL:X57993; NID:950476; PIDN:CAA41048.1; PID:950477
C:Genetics:
A:Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 2
58/3; 673/3; 706/3; 742/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix
F:1-24/Domain: signal sequence status predicted <SIG>
F:25-154/Domain: propetide status predicted <PRO>
F:32-92/Domain: von Willebrand factor type C repeat homology <WVC>
F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>
F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 53.3%; Score 1857.5; DB 2; Length 1464;
Best Local Similarity 53.4%; Pred. No. 1.6e-86;
Matches 369; Conservative 39; Mismatches 190; Indels 93; Gaps 12;
QY 1 GPPGEPGTGLGPPGGERGGSGRFGPCADGVAGPKGAPAGERGSGPAGPKGSGPAGR 60
DB 467 GSPGEPGANGLPGAAGRGSGFRGPAFNGIPGKPPGPPGPPGPPGPPGPPGPPG 526
QY 61 GAGLPGAKLTGSGSPGPDGKTGPPGAGDGRPPGPPGPPGPPGPPGPPGPPGPPG 120
DB 527 GTPGFGIRGMPGSGPPGNDGKPPGSGSGSPGPPGPPGPPGPPGPPGPPGPPG 566
QY 121 GPPGKAGERGVPPGPAVGPAGKGEAGAGQPPGAPGAPGAPGAPGAPGAPGAPG 180
DB 587 GAPKNGERGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 646
QY 181 GPPGKAGEGQGVPPDGLGAPGSPGAPGPPGTGLPFPFPPGPPGPPGPPGPPGPP 240
DB 647 GPPGKAGEGQGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 700
QY 241 GPAGERSGPPGAPGKSGPGEAGRPGGAGLPGAKLTGSGSPGPPGPPGPPGPPG 300
DB 701 GP---EGGKGPAGPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 754
QY 301 GPPGPPGARGQAGVMGFPGPKGAAGPGKAGRGVPPGPAVGPAGKGEAGAGQPPG 360
DB 755 GNDGP---RGAPGPPGPP---AGQPKGSGSGPGLPGIAGPRGPPGPPGPPGPP 808
QY 361 -----GPAGERGQPPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 408
DB 809 GPPGAPQNGEPGKAGRGAPGKGGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 868
QY 409 GPAGEGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 468
DB 869 GTAGFPGGRLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 928

Am. J. Hum. Genet. 53, 62-70, 1993

A/Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual splicing.

A/Reference number: I51868; MUID:93304430; PMID:8317500

A/Accession: I51868

A/Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: DNA

A/Residues: 186-194 <ML>

A/Cross-references: GB:S62925; NID:G386425; PIDN:AAD13937.1; PID:G4261637

R/Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.

Biochem. J. 311, 939-943, 1995

A/Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A1 gene.

A/Reference number: S59511; MUID:96087614; PMID:7487954

A/Accession: S59511

A/Molecule type: mRNA

A/Residues: 302-423 <CH>

A/Cross-references: GB:S79877; NID:G1195576; PIDN:AAB35615.1; PID:G1195577

R/Seyer, J.M.; Kang, A.H.

Biochemistry 17, 3404-3411, 1978

A/Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr peptides.

A/Reference number: A90414; MUID:79000343; PMID:687591

A/Accession: A90414

A/Molecule type: protein

A/Residues: 399-675, 'N', 677-727 <SEY3>

A/Experimental source: liver

R/Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.

J. Biol. Chem. 266, 5256-5259, 1991

A/Title: G to T transversion at position +5 of a splice donor site causes skipping of exon 17.

A/Reference number: I55349; MUID:91161621; PMID:1672129

A/Accession: I55349

A/Status: translated from GB/EMBL/DBD

A/Molecule type: DNA

A/Residues: 537-605 <LEE>

A/Cross-references: GB:M59312; NID:G180815; PIDN:AAA52041.1; PID:G180816

R/Seyer, J.M.; Mainardi, C.; Kang, A.H.

Biochemistry 19, 1583-1589, 1980

A/Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from human skin.

A/Reference number: A90438; MUID:80198282; PMID:6246925

A/Accession: A90438

A/Molecule type: protein

A/Residues: 728-835, 'A', 897-964 <SEY4>

A/Experimental source: liver

R/Cole, W.G.; Chiodo, A.A.; Llanande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chaudhary, K.

J. Biol. Chem. 265, 17070-17077, 1990

A/Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping.

A/Reference number: A38303; MUID:91009133; PMID:2145268

A/Accession: A38303

A/Molecule type: mRNA

A/Residues: 861-1015 <COL>

A/Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AAB59383.1; PID:J05617

A/Note: A mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syndrome.

R/Mankoo, B.S.; Dalgleish, R.

Nucleic Acids Res. 16, 2337, 1988

A/Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.

A/Reference number: S02119; MUID:88189827; PMID:3357782

A/Accession: S02119

A/Status: translation not shown

A/Molecule type: mRNA

A/Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>

A/Cross-references: EMBL:X06700; NID:G30053; PIDN:CAA29886.1; PID:G30054

R/Seyer, J.M.; Kang, A.H.

Biochemistry 20, 2621-2627, 1981

A/Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from human skin.

A/Reference number: A90446; MUID:81208139; PMID:7016180

A/Accession: A90446

A/Molecule type: protein

A/Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157

A/Experimental source: liver

R/Lodl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Myllyluoma, J.

Nucleic Acids Res. 12, 9383-9394, 1984

A/Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollagen.

A/Reference number: A93551; MUID:85087944; PMID:6096827

A/Accession: A93551

A:Molecule type: mRNA
 A:Residues: 1065-1155, 'P', 1157-1466 <LOI>
 A:Cross-references: EMBL:X01655; EMBL:X01742; NID:G29584; PIDN:CAA25821.1
 R:Miskulin, M.; Daigleish, R.; Klueve-Beckerman, B.; Renard, S.I.; Tolstoshev, P.; Brant
 Biochemistry 25, 1408-1413, 1986
 A>Title: Human type III collagen gene expression is coordinately modulated with the type
 A:Reference number: 152393; MUID:86187804; PMID:3754462
 A:Accession: I52393
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1161-1200 <MIS>
 A:Cross-references: GB:M13146; NID:G180415; PIDN:AAA52003.1; PID:G180416
 R:Emanuel, B.S.; Cannizzaro, L.A.; Seiver, J.M.; Myers, J.C.
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
 A>Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
 A:Reference number: 159025; MUID:85216505; PMID:3858826
 A:Accession: I79359
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1165-1196 <EMA>
 A:Cross-references: GB:M11134; NID:G180417; PIDN:AAA52004.1; PID:G180418
 R:Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sipola, M.; Ramirez, F.
 J. Biol. Chem. 260, 4357-4363, 1985
 A>Title: Isolation of cDNA and genomic clones encoding human pro-alpha(III) collagen.
 A:Reference number: A92516; MUID:85157600; PMID:2579949
 A:Accession: A92516
 A:Molecule type: DNA
 A:Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>
 A:Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB
 A:Experimental source: liver
 A>Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given
 action
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently
 C:Genetics:
 A:Gene: GDB:COL3A1
 A:Cross-references: GDB:118729; OMIM:120180
 A:Map position: 2q31-2q31
 A:Introns: 27/; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
 A>Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan
 C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfide
 er of their length, is formed with desmosine cross-links made from lysine and allylsine
 C:Function:
 A:Description: structural component of extracellular fibrous polymer that maintains inte
 C:Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyc
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-153/Domain: amino-terminal propeptide #status predicted <PRO>
 F:31-91/Domain: von Willebrand factor type C repeat homology <VWC>
 F:154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>
 F:154-167/Region: amino-terminal nonhelical telopeptide
 F:168-1196/Region: helical
 F:1091-1093/Region: cell attachment (R-G-D) motif
 F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide
 F:1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
 F:1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 F:24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F:153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted
 F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F:163-1612/Modified site: allylsine (Lys) #status predicted
 F:263-284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:263/Binding site: carboxylate (Lys) (covalent) #status experimental
 F:584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
 F:948-949/Cleavage site: Gly-Ile (collagenase) #status experimental
 F:1106/Binding site: carboxylate (Lys) (covalent) #status predicted

Query Match

Best Local Similarity 53.1%; Score 1853; DB 1; Length 1466;

C:Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;

Matches 365; Conservative 42; Mismatches 209; Indels 108; Gaps 11;

Qy 1 GPPGEPGTGLPGRGGGSGRGGGADGVAGKPGAGRGSPGAPGKSPGAGRP 60

Db 468 GSPGEPGANGLPAGAGRGAGFGFGAGNGIPGKPGAGRGAPGAPGAGRGAGRG 527

Qy 61 GEAGLPKAGLGTSGSPGSPGDPGKTPGPGPAGODGRPGPPGARGOAGVGMFFGPKGAA 120
 Db 528 GVFGGDMGMPGSPGSGSDGKPGPPGSGSGRPPGPPGSGGPPGQGVGMFFGPKGND 587
 Qy 121 GSPGKAGRGVPPGAVGAFAGKDEAGAAQAGPFPAGPAGRGEGQAGSPAGSPGQGLGPA 180
 Db 588 GAFKNGRGSGGPGPGPQCPGPKNGETGPGQPPGTGPGDKGDTGPPGQGLGQLGPGT 647
 Qy 181 GPPGKAGRGEGVPGDLAGP-----GPSGAPGEP 210
 Db 648 GPPGKNGKPGEPGPKGDAGAPAGPKGDAGAPGKGDAGAPGPPGLAGAPGLRGAGGAPGPEGK 707
 Qy 211 GPTGLPFPGRGPGSGRGFPAGDGVAGPKGAPGERSP-----GPAGPK 255
 Db 708 GAAGPFPGSAAGTGLQMPGERGLSPGPKDKGPGGADGVPGKDPGPGTGP 767
 Qy 256 GSPGAGRP---GEAGLPKAGLGTSGSPGSPDGTGTPPGAGQDGRGPPGPPGARGQA 312
 Db 768 GPPGAGQPDGKGGGAPGLFGIAGPRGSPGERGTGTPPGAGFPAGPQNGEPGKGER 827
 Qy 313 GV---MGFPGRKGAAGEPKKAGRGVGP-----PGAVGPAKDGCEA 351
 Db 828 GAPGKGGGPPGAVAGPPGGSPAGPDPQGVKGRGSPGPGAGFPAGRLGPPGPN 887
 Qy 352 GAQGPFPAGPAGRGEGQAGSPGFPGLPAGPAGPAGKPGEGVPGDLAGPAGPAG 411
 Db 888 GNPFPSPGSGCKDGPFGAGTGAAGSPGVPAGKDGAGQPGKSGSPAGPAGPAG 947
 Qy 412 GPGTGL-----PGPGRGSGPGRGCF-----PGADGVAGKPGAGRGSPGAPGK 459
 Db 948 GIAGTGTAGLGGPMPGPPGSGPQGVKSGKPGANGLSGRGPPGPGGLPGLAGTA 1007
 Qy 460 GSPGAGRPGEAGLPGA-----KGLTSGSPGSPDGTGTPPGP-----AGQDGRP 504
 Db 1008 GEPGRDGNFGSDGLPGRDGSFGKGRGNGSPGAPGAPGHPGPPGVPAGKSGDRGES 1067
 Qy 505 GPPGFPARGQAGVGMFFGPKGAGEPKAGRGV-----PGPPGAVGPAKDGCEA 555
 Db 1068 GPAGPAGPAGPAGRGAGPAGPQGPGRGDKGTGERGAAGIKHGRGPPGPPGAPGSPGAGQ 1127
 Qy 556 GAQGPFPAGPAGRGEGQAGSPGFPGLPAGPAGPAGKPGEGVPGDLAGPAGPAG 612
 Db 1128 GAIGSGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 1187
 Qy 613 GPAG 616
 Db 1188 GPPG 1191

RESULT 15

I50694

collagen alpha 1(III) chain - chicken (fragment)

C:Species: Gallus gallus (chicken)

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004

C:Accession: I50694

R:Nah, H.D.; Niu, Z.; Adams, S.L., 1994

J. Biol. Chem. 269, 16443-16448, 1994

A>Title: An alternative transcript of the chick type III collagen gene that does not enc

A:Reference number: A54041; MUID:94266842; PMID:8206952

A:Accession: I50694

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-886 <NAH>

A:Cross-references: UNIPROT:P12105; EMBL:U07973; NID:G520454; PIDN:AAA83407.1; PID:G5374

C:Genetics:

A:Gene: COL3A1

C:Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;

F:30-90/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 52.8%; Score 1840; DB 2; Length 886;

Best Local Similarity 49.8%; Pred. No. 8.4e-86;

Matches 362; Conservative 44; Mismatches 195; Indels 126; Gaps 11;

Search completed: October 18, 2004, 13:42:52
Job time : 29.4569 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:23:10 ; Search time 110.271 Seconds
(without alignments)
3219.411 Million cell updates/sec

Title: US-10-658-989A-2

Perfect score: 3488

Sequence: 1 GPPGCPGTLGPPGPRGG.....GEGVPGDLGAPGSPGAGG 617

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 57537466 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2237.5	64.1	1464	1 CAIL1_HUMAN	P02452 homo sapien
2	2233.5	64.0	1461	2 Q76045	Q76045 homo sapien
3	2233.5	64.0	1464	2 Q8N473	Q8N473 homo sapien
4	2231.5	64.0	1460	1 CAIL1_CANFA	Q9XSJ7 canis fami
5	2222	63.7	1069	2 Q6LAN8	Q6LAN8 homo sapien
6	2222	63.7	1069	2 CAA67261	CAa67261 homo sapi
7	2215.5	63.5	1453	2 Q63079	Q63079 rattus norv
8	2205.5	63.2	1433	1 CAIL1_MOUSE	P11087 mus musculu
9	2205.5	63.2	1453	2 Q81009	Q81009 mus musculu
10	2176.5	62.4	1225	2 Q6PCU3	Q6PCU3 mus musculu
11	2176.5	62.4	1225	2 AAH59281	AAH59281 mus muscu
12	2151.5	61.7	1453	1 CAIL1_CHICK	P02457 gallus gall
13	2043.5	58.6	1450	2 Q9YIB4	Q9YIB4 cynops pyrr
14	2043.5	58.6	1445	2 Q93251	Q93251 rana catesb
15	2029	58.2	1449	2 Q80285	Q80285 xenopus lae
16	2010.5	57.6	671	1 CAIL1_RAT	P02454 rattus norv
17	1976.5	56.7	1447	2 Q9IB91	Q9IB91 xenopus lae
18	1953	56.0	1160	2 Q14046	Q14046 homo sapien
19	1953	56.0	1487	2 Q14047	Q14047 homo sapien
20	1952	56.0	1487	2 Q77753	Q77753 canis fami
21	1944	55.7	1418	2 Q28396	Q28396 equus cabal
22	1943	55.7	1418	1 CAIL2_HUMAN	P02458 homo sapien
23	1942	55.7	1420	2 Q90W37	Q90W37 gallus gall
24	1937	55.5	1269	2 Q7T227	Q7T227 mus musculu
25	1936	55.5	1442	2 Q62031	Q62031 mus musculu
26	1936	55.5	1442	2 Q62033	Q62033 mus musculu
27	1936	55.5	1459	1 CAIL2_MOUSE	P28481 mus musculu
28	1936	55.5	1459	2 Q62032	Q62032 mus musculu
29	1927	55.2	1419	2 Q80VX3	Q80VX3 mus musculu
30	1927	55.2	1419	2 Q80X38	Q80X38 mus musculu
31	1923	55.1	1447	2 Q6ULU5	Q6ULU5 brachydanio

32 1923 55.1 1447 2 AAR24536
33 1921 55.1 1419 2 Q63123
34 1914 54.9 1447 2 Q6P4UI
35 1914 54.9 1447 2 AAh63249
36 1898 54.4 779 1 CAIL1_BOVIN
37 1892 54.2 1492 2 Q6P4Z2
38 1892 54.2 1486 2 AAh63191
39 1890.5 54.2 1486 2 Q91717
40 1888.5 54.1 1486 2 Q72TI6
41 1883 54.0 1418 2 Q9W7R9
42 1880 53.9 1491 2 Q7ZTM4
43 1880 53.9 1491 2 Q91718
44 1873 53.7 1262 1 CAIL3_CHICK
45 1873 53.7 1449 2 Q910C0

ALIGNMENTS

RESULT 1
Call_HUMAN STANDARD; PRT: 1464 AA.
ID AC P02452; P78441; Q13896; Q13902; Q13903; Q14037; Q14992; Q15176;
AC Q15201; Q16050; Q7KZ30; Q7KZ34; Q8IV15; Q9UML6; Q9UMW7;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT SER-1434.
RA Dalgleish R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBSJ databases.
RN [2]
RP SEQUENCE OF 1-589 FROM N.A.
RX PubMed=2843432;
RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,
RA Pretorius P.J.;
RT "Complete nucleotide sequence of the region encompassing the first
twenty-five exons of the human pro alpha 1(I) collagen gene
(COL1A1).";
RL Gene 67:103-115 (1988).
RN [3]
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
JAenisch R., Prockup D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
of human type I procollagen.";
RL Biochem. J. 253:919-922 (1988).
RN [4]
RP SEQUENCE OF 1-181 FROM N.A.
RX MEDLINE=84720697; PubMed=6462220;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
Myers J., Williams C., Ramirez F.;
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
conservation of a pattern of introns and exons.";
RL Nature 310:337-340 (1984).
RN [5]
RP SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.
RC TISSUE=Skin;
RX MEDLINE=71038625; PubMed=5529814;
RA Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
the alpha 1 and alpha 2 chains of human skin collagen.";
RL Biochemistry 9:4699-4706 (1970).
RN [6]
RP SEQUENCE OF 425-1464 FROM N.A.
RX MEDLINE=84080385; PubMed=6689127;

- RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
 RA Prockop D.J.;
 RA "Nucleotide sequences of complementary deoxyribonucleic acids for the
 RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
 RT of structures that are conserved during evolution.";
 RL Biochemistry 22:5213-5223(1983).
 RN [7]
 RP SEQUENCE OF 472-607 FROM N.A.
 RX PubMed=2981843;
 RA Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;
 RA "Multixon deletion in an osteogenesis imperfecta variant with
 RT increased type III collagen mRNA.";
 RL J. Biol. Chem. 260:691-694(1985).
 RN [8]
 RP SEQUENCE OF 488-625 FROM N.A.
 RX PubMed=3857621;
 RA Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;
 RA "Intron-mediated recombination may cause a deletion in an alpha 1 type
 RT I collagen chain in a lethal form of osteogenesis imperfecta.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874(1985).
 RN [9]
 RP SEQUENCE OF 1179-1464 FROM N.A., VARIANTS OI-II H1S-1277; ARG-1388 AND
 RP 1337-GLU-1338 DEL, AND VARIANT SER-1434.
 RX MEDLINE=93352646; PubMed=8349697;
 RA Chessler S.D., Wallis G.A., Byers P.H.;
 RA "Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)
 RT chain of type I collagen result in defective chain association and
 RT produce lethal osteogenesis imperfecta.";
 RL J. Biol. Chem. 268:18218-18225(1993).
 RN [10]
 RP SEQUENCE OF 1229-1454 FROM N.A.
 RC TISSUE=bone;
 RX MEDLINE=88124208; PubMed=3340531;
 RA Maekelae J.K., Raessina M., Virts A., Vuorio E.;
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 RT domain.";
 RL Nucleic Acids Res. 16:349-349(1988).
 RN [11]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=85130970; PubMed=2857713;
 RA Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
 RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
 RT Promoter structure, Alu repeats, and polymorphic transcripts.";
 RL J. Biol. Chem. 260:2315-2320(1985).
 RN [12]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=88097389; PubMed=3480516;
 RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
 RT "Regulatory elements in the first intron contribute to transcriptional
 RT control of the human alpha 1(I) collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
 RN [13]
 RP SEQUENCE OF 1-44 FROM N.A.
 RX MEDLINE=88033098; PubMed=2822714;
 RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
 RA de Wet W.;
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)
 RT collagen gene enhance transcription.";
 RL J. Biol. Chem. 262:15151-15157(1987).
 RN [14]
 RP SEQUENCE OF 175-187 AND 274-289.
 RX PubMed=2169412;
 RA Baetge B., Notbohm H., Diebold J., Lehmann H., Bodo M., Deutzmann R.,
 RA Muller P.K.;
 RT "A critical crosslink region in human-bone-derived collagen type I.
 RT Specific cleavage site at residue Leu95.";
 RL Eur. J. Biochem. 192:153-159(1990).
 RN [15]
 RP SEQUENCE OF 263-268.
 RC TISSUE=skin;
 RX MEDLINE=71001508; PubMed=4319110;
 RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
 RT "A comparative study of glycopeptides derived from selected vertebrate
 RT collagens. A possible role of the carbonyl group in fibril formation.";
 RL J. Biol. Chem. 245:5042-5048(1970).
 RN [16]
 RP SEQUENCE OF 281-302; 402-420; 823-842; 924-944; 1026-1045 AND
 RP 1143-1162 FROM N.A.
 RX PubMed=2374517;
 RA Labhard M.E., Hollister D.W.;
 RT "Segmental amplification of the entire helical and telopeptide regions
 RT of the cDNA for human alpha 1 (I) collagen.";
 RL Matrix 10:124-130(1990).
 RN [17]
 RP SEQUENCE OF 425-490; 965-1024; 999-1039 AND 1453-1464 FROM N.A.
 RX MEDLINE=83064528; PubMed=6183642;
 RA Chu M.-L., Myers J.C., Bernard M.P., Ding J.-F., Ramirez F.;
 RT "Cloning and characterization of five overlapping cDNAs specific for
 RT the human pro alpha 1(I) collagen chain.";
 RL Nucleic Acids Res. 10:5925-5934(1982).
 RN [18]
 RP SEQUENCE OF 710-745 FROM N.A., AND VARIANT OI-II ARG-728.
 RX PubMed=2339700;
 RA Wallis G.A., Starman B.J., Zinn A.B., Byers P.H.;
 RT "Variable expression of osteogenesis imperfecta in a nuclear family is
 RT explained by somatic mosaicism for a lethal point mutation in the
 RT alpha 1(I) gene (COL1A1) of type I collagen in a parent.";
 RL Am. J. Hum. Genet. 46:1034-1040(1990).
 RN [19]
 RP SEQUENCE OF 746-781 FROM N.A., AND VARIANT OI-III SER-767.
 RX MEDLINE=95187161; PubMed=7881420;
 RA Forlino A., Zolezzi F., Valli M., Pignatti P.F., Cetta G.,
 RA Brunelli P.C., Mettes M.;
 RT "Severe (type III) osteogenesis imperfecta due to glycine
 RT substitutions in the central domain of the collagen triple helix.";
 RL Hum. Mol. Genet. 3:2201-2206(1994).
 RN [20]
 RP SEQUENCE OF 1187-1220 FROM N.A., AND VARIANT CVS-1195.
 RX PubMed=3170557;
 RA Cohn D.H., Apone S., Eyre D.R., Starman B.J., Andressen P.,
 RA Charbonneau H., Nicholls A.C., Pope F.M., Byers P.H.;
 RT "Substitution of cysteine for glycine within the carboxyl-terminal
 RT telopeptide of the alpha 1 chain of type I collagen produces mild
 RT osteogenesis imperfecta.";
 RL J. Biol. Chem. 263:14605-14607(1988).
 RN [21]
 RP SEQUENCE OF 1440-1454 FROM N.A.
 RX MEDLINE=90110450; PubMed=2295701;
 RA Willing M.C., Cohn D.H., Byers P.H.;
 RT "Frameshift mutation near the 3' end of the COL1A1 gene of type I
 RT collagen predicts an elongated pro alpha 1(I) chain and results in
 RT osteogenesis imperfecta type I.";
 RL J. Clin. Invest. 85:282-290(1990).
 RN [22]
 RP SEQUENCE OF 1454-1464 FROM N.A.
 RX PubMed=1995349;
 RA Maatta A., Bornstein P., Penttinen R.P.;
 RT "Highly conserved sequences in the 3'-untranslated region of the
 RT COL1A1 gene bind cell-specific nuclear proteins.";
 RL FEBS Lett. 279:9-13(1991).
 RN [23]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=91184577; PubMed=2010058;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in collagen genes: causes of rare and some common diseases
 RT in humans.";
 RL FASEB J. 5:2052-2060(1991).
 RN [24]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 RN [25]

```
RP REVIEW ON OI VARIANTS.
RX MEDLINE=91374476; PubMed=1895312;

Query Match      64.1%; Score 2237.5; DB 1; Length 1464;
Best Local Similarity 60.2%; Pred. No. 9.5e-83;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

Qy 1 GPP-----GSPGTGTLPGPPGRRGGSGFPDAGVAGPKGPPGSGP 48
Db 458 GPPGAGGKRGAGSGPPTGLPPGRRGGSGFPDAGVAGPKGPPGSGP 517
Qy 49 GPKSGPGEAGPAGAGLPGAKGLTGTSGSPGPPGKGTGPPGAGDGRPPGPPGARGQA 108
Db 518 GPKSGPGEAGPAGAGLPGAKGLTGTSGSPGPPGKGTGPPGAGDGRPPGPPGARGQA 577
Qy 109 GVMGFPGPKGAAGPFGKAGRGVPGPPGAVGPAKDGCEAGCAQPPGPPGAGRGEGGPA 168
Db 578 GVMGFPGPKGAAGPFGKAGRGVPGPPGAVGPAKDGCEAGCAQPPGPPGAGRGEGGPA 637
Qy 169 GSPGFGPLGPPAGPAGKAGRGVPGPPGAVGPAKDGCEAGCAQPPGPPGAGRGEGGPA 212
Db 638 GSPGFGPLGPPAGPAGKAGRGVPGPPGAVGPAKDGCEAGCAQPPGPPGAGRGEGGPA 697
Qy 213 -----TGLPGPPGRRGGSGFPDAGVAGPKGPPGSGPAGPKGSP 258
Db 698 GAGAPGNDGAKGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 757
Qy 259 GE-----AGPGEAGLPGAK-----GLTSGPSPGPPGKGTGPPGAGDGRPP 300
Db 758 GKDGVRLGTGTLGPPGPPGAGPAGKAGRGVPGPPGAVGPAKDGCEAGCAQPPGPPGAGP 817
Qy 301 GPPGPPGARGO-----AGVVGFPKGKAGRGVPGPPGAGPAGPKGSP 341
Db 818 GADQPAKAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 877
Qy 342 -----VGPAGKDGAGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 393
Db 878 GAAGRVPPGPGSGNAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 937
Qy 394 GEQGVGDLNAGPSPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 453
Db 938 GADGAPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 997
Qy 454 GPAGP---KGSFGBAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 510
Db 998 GPMGPPGLAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1057
Qy 511 GARGQAGVMGPPGPKGAAGPFGKAGRGVPGPPGAVGPAKDGCE-----AGAQ 558
Db 1058 GPAGKSGDRGTGTGAPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1117
Qy 559 GPPGP-----AGPAGRGEGGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 603
Db 1118 GPPGPPGSGGSGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1177
Qy 604 GDLGAPGSPGAG 616
Db 1178 GPPGPPGPPGPPG 1190

RESULT 2
ID 076045 PRELIMINARY; PRT; 1461 AA.
AC 076045;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Pro alpha 1(I) collagen.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, Alu repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
[2]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=88329734; PubMed=2843432;
RA D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene.";
RL Gene 67:1105-115(1988).
[3]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
[4]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=91138770; PubMed=1995349;
RA Maatta A., Bornstein P., Penttinen R.P.;
RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";
RL FEBS Lett. 279:9-13(1991).
[5]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=92157916; PubMed=1787829;
RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL Matrix 11:375-379(1991).
[6]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=98107942; PubMed=9443882;
RA Korkko J.M., Earley J.J., Nuytinck L., De Paape A., Prockop D.J.,
RA Ala-Kokko L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017178; AAB94054.2; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0003737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PSS0184; VWF_C; 1.
KW Collagen.
SQ SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30EA78E21 CRC64;
Query Match      64.0%; Score 2233.5; DB 2; Length 1461;
```

```

Best Local Similarity 60.0%; Pred. No. 1.4e-82;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGERGSGRFPFGADGVAGPKGAGRGSGPGA 48
DB 455 GPPGAGBEGKRGARGEPGPTGLPGERGSGRFPFGADGVAGPKGAGRGSGPGA 514
QY 49 GPKSGPAGRPGEAGLFGAKGLTSGPSGSPGDKTGTTPPGAGODGRRPFPFGARGQA 108
DB 515 GPKSGPAGRPGEAGLFGAKGLTSGPSGSPGDKTGTTPPGAGODGRRPFPFGARGQA 574
QY 109 GVMGFPKGAGBPKKAGRGVPPGCAVCPAGKDGAGAGNQGPPGAGPAGERGEGPA 168
DB 575 GVMGFPKGAGBPKKAGRGVPPGCAVCPAGKDGAGAGNQGPPGAGPAGERGEGPA 634
QY 169 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLCAPGSPGAGE-----PGP--- 212
DB 635 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLCAPGSPGAGE-----PGP--- 212
QY 213 -----TGLPFPGERGSGRFPFGADGVAGPKGAGRGSGPGA 694
DB 695 GAGAPGNDGAKGADGAGPAGFSGQAGFLQMPGGERGAAGLPGPKGRDAGPKGADGSP 754
QY 259 GE-----AGRPGEAGLFGAK-----GLTSGPSGSPGDKTGTTPPGAGODGRRP 300
DB 755 GKDVGRLTGLPFPAGAPGDDTGESGSPGAPGPTGARGAPGDRGEPFGPPGAGPAGPP 814
QY 301 GPPGPPGARGQ-----AGVMFPKPKGAAGEPGKAGRGVPPGPGA----- 341
DB 815 GADGQPKAGBPGDAGAKGADGAPGAPGPPGPIGNVAPKAGARGAGPAGATGFP 874
QY 342 -----VGPAGKDGAGAGPAGPAGP-----AGERGQAGPAGSPGQGLPAGPAGK 393
DB 875 GAARVGPFGSGNAGPPGPPGAKGKGKGRGETGAPRGVGPFGPPGAGEKSP 934
QY 394 GEGVPGDLGAPGSPGAGEPGLPGERGSGRFPFGADGVAGPKGAGRGSGP 453
DB 935 GADGAPAGPPTGPGIAGQGVVGLPGQGERGFPGLPSPGEPKQKSGASGERGPP 994
QY 454 GPAGP-----KSPGAGRPGEAGLFGAKGLTSGPSGSPGDKTGTTPPGAGODGRRP 510
DB 995 GPMGPPGLAGPFGSGREGAPGAGSPGDRGTPGAGPAGPAGPAGPAGPV 1054
QY 511 GARGQAGVMGFPKGAAGEPKAGRGVPPGCAVCPAGKDGAGAGNQGPPGAGPAGERG 558
DB 1055 GPAGKSGDRGTGTGAPGAPGVPGVARGAPGQPGDKGTGEGQGRGKGRGSGLO 1114
QY 559 GPPGP-----AGPAGERGEGPAGSP-----GFOGLPAGPAGPAGKRGSGV 603
DB 1115 GPPGPPGSPGQPSGASGAPRGPPGSGAGPKDGLNGLPGPIGPPGRGRTGAGPV 1174
QY 604 GDLGAPGSPGAG 616
DB 1175 GPPGPPGPPGPPG 1187

RESULT 3
Q8N473 PRELIMINARY; PRT; 1464 AA.
AC Q8N473;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen, preproprotein.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22389257; PubMed=12477932;

```

```

Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buotow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Scapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahy J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Souffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodríguez A.C., Grinwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
Krzyszewski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
Jones S.J., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
SEQUENCE FROM N.A.
TISSUE=Brain;
Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
EMBL; BC036533; AAH36531.1; -
GO; GO:0005581; C:cytoplasm; IEA.
GO; GO:0005737; C:cytoplasm; IEA.
GO; GO:0006817; P:phosphate transport; IEA.
InterPro; IPR008161; Cig helix.
InterPro; IPR008160; Collagen.
InterPro; IPR000885; Fib collagen_C.
InterPro; IPR009041; PMP_SGCI.
InterPro; IPR001007; VWF_C.
Pfam; PF01410; COLFI; 1.
Pfam; PF01391; Collagen; 18.
ProDom; PD000007; Cig helix; 3.
ProDom; PD002078; Fib collagen_C; 1.
SMART; SM00038; COLFI; 1.
SMART; SM00214; VWC; 1.
PROSITE; PS01208; VWF_C_1; 1.
PROSITE; PS0184; VWF_C_2; 1.
Collagen.
KW
SQ
SEQUENCE 1464 AA; 139010 MW; B0581F8D1C89DDE8 CRC64;

Query Match 64.0%; Score 2233.5; DB 2; Length 1464;
Best Local Similarity 60.0%; Pred. No. 1.4e-82;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGERGSGRFPFGADGVAGPKGAGRGSGPGA 48
DB 458 GPPGAGBEGKRGARGEPGPTGLPGERGSGRFPFGADGVAGPKGAGRGSGPGA 517
QY 49 GPKSGPAGRPGEAGLFGAKGLTSGPSGSPGDKTGTTPPGAGODGRRPFPFGARGQA 108
DB 518 GPKSGPAGRPGEAGLFGAKGLTSGPSGSPGDKTGTTPPGAGODGRRPFPFGARGQA 577
QY 109 GVMGFPKGAGBPKKAGRGVPPGCAVCPAGKDGAGAGNQGPPGAGPAGERGEGPA 168
DB 578 GVMGFPKGAGBPKKAGRGVPPGCAVCPAGKDGAGAGNQGPPGAGPAGERGEGPA 637
QY 169 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLCAPGSPGAGE-----PGP--- 212
DB 638 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLCAPGSPGAGE-----PGP--- 212
QY 213 -----TGLPFPGERGSGRFPFGADGVAGPKGAGRGSGPGA 697
DB 698 GAGAPGNDGAKGADGAGPAGFSGQAGFLQMPGGERGAAGLPGPKGRDAGPKGADGSP 757
QY 259 GE-----AGRPGEAGLFGAK-----GLTSGPSGSPGDKTGTTPPGAGODGRRP 300
DB 758 GKDVGRLTGLPFPAGAPGDDTGESGSPGAPGPTGARGAPGDRGEPFGPPGAGPAGPP 817

```

301 GPPPPGARGQ-----AGVMGFPCKAAGPGKAGRGVPGPPGA----- 341.
818 GADGCPGAKGPGDAGAKGDAGPPGAPGPPGICNVGAPGAKGARGAGPPGATGFP 877
342 -----VGPAGKDGAGAGQGGPPGAPG---AGERGEQPGAPSPGQGLPGPAGPPGAGK 393
878 GAAGRVGPPGSPGNAGPPGPPGAGKGGKRGRTGPGAGPGEVGGPPGPPGAGKGGSP 937
394 GEQGVPGDLGAPGSPGAPGPPGTGLPGPGERGPGSRGPPGADGVAGPKGAPAGRGSP 453
938 GADGPAGAGTGGPQGIAGORGVVGLPGQGRGPPGLPGPSPGPPGKQSPGSGASGPP 997
454 GPAGP---KSGPGEAGRPGAGLPGAKGLTGSPGSPGPDGKTGTPPGAGQDGRPPGPP 510
998 GPMGPPGLAGPPGSGREGAPGARGSPGDRGSPGAKGDRGTGPGAGPAGPAGPAGPV 1057
511 GARGQAGVMGPPGKGAAGPGKAGRGVPGPPGAVGAPGKDGPE-----AGAQ 558
1058 GPAGKSGDRGTGPGAGPAGPVGARGPAGPQGRGDKGTGEGQDRGKIGHRGFSLG 1117
559 GPPGPP-----AGPAGERGEGPAGSP---GFQGLPGPAGPPGAGKGGVGP 603
1118 GPPGPPGSPGQSPGSGAGPPGPPGSPGAPGKQGLNGLPGPPGPPGRTGDRGPV 1177
604 GDLGAPGSPGAP 616
1178 GPPGPPGPPGPPG 1190

RESULT 4
CALL CANFA
ID CALL CANFA STANDARD; PRT; 1460 AA.
AC Q9XSJ7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DI 01-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OM Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxId=9615;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT OF ALA-208.
RC TISSUE=Skin;
RX MEDLINE=21023337; PubMed=11147834;
RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;
RT "Sequence of normal canine COL1A1 cDNA and identification of a
RT heterozygous alpha1(I) collagen Gly208Ala mutation in a severe case of
RT canine osteogenesis imperfecta.";
RL Arch. Biochem. Biophys. 384:37-46(2000).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta
CC (OI).
CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
CC -!- SIMILARITY: Contains 1 VWFC domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF153062; AAD34619.1; -;
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.

DR InterPro; IPR00885; Fib collagen_C.
DR InterPro; IPR001007; VWFC.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; C1g_helix; 2.
DR ProDom; PD002078; Fib collagen_C; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS0184; VWFC_2; 1.
KW Collagen; Connective tissue; Disease mutation; Extracellular matrix;
KW Glycoprotein; Hydroxylation; Pyrrolidone carboxylic acid; Repeat;
KW Signal.
FT SIGNAL 1 22 By similarity.
FT PROPEP 23 157 N-terminal propeptide.
FT CHAIN 158 1214 Collagen alpha 1(I) chain.
FT PROPEP 1215 1460 C-terminal propeptide.
FT DOMAIN 34 92 VWFC.
FT DOMAIN 158 174 Nonhelical region (N-terminal).
FT DOMAIN 175 1188 tripe-helical region.
FT DOMAIN 1189 1214 Nonhelical region (C-terminal).
FT SITE 741 743 Cell attachment site (potential).
FT SITE 1089 1091 Cell attachment site (potential).
FT MOD_RES 158 158 Pyrrolidone carboxylic acid (By similarity).
FT MOD_RES 166 166 Allylsine (By similarity).
FT MOD_RES 261 261 S-hydroxylysine (By similarity).
FT CARBOHYD 1160 1160 3-hydroxyproline (By similarity).
FT CARBOHYD 261 261 O-linked (Gal..) (By similarity).
FT CARBOHYD 1361 1361 N-linked (GlcNAc..) (By similarity).
FT VARIANT 208 208 G -> A (in OI; severe).
SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;
Query Match 64.0%; Score 2231.5; DB 1; Length 1460;
Best Local Similarity 60.2%; Pred. No. 1.7e-82;
Matches 441; Conservative 20; Mismatches 155; Indels 117; Gaps 12;
QY 1 GPP-----GEPPTGLPQPPQGERGPPGSRGPPGADGVAGPKGAPGERSGPA 48
Db 454 GPPGAGEGKRGAGEPQPTGLPQPPGGERGPPGSRGPPGADGVAGPKGAPGERSGPA 513
QY 49 GPKGSPGEAGRPCEAGLPGAKGLTGSPSPGPDGKTGTPPGAGQDGRPPGPPGARGQA 108
Db 514 GPKGSPGEAGRPCEAGLPGAKGLTGSPSPGPDGKTGTPPGAGQDGRPPGPPGARGQA 573
QY 109 GVMGPPGKGAAGEPKGAGERGPPGAVGAGKDGAGAGOGPPGAPGAPGERGEGQA 168
Db 574 GVMGPPGKGAAGEPKGAGERGPPGAVGAGKDGAGAGOGPPGAPGAPGERGEGQA 633
QY 169 GSPGFQGLPGAPPPGAEAGKPGEQGVPGDLGAPSPGAGE-----PGP----- 212
Db 634 GSPGFQGLPGAPPPGAEAGKPGEQGVPGDLGAPSPGAGERGPPGAGPPGPPGAPGR 693
QY 213 -----TGLPQPPGERGPPSRGPPGADGVAGPKGAPGERSGPPGAPGKSP 258
Db 694 GAGAPGNDGAKGDAGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 753
QY 259 GE-----AGRPCEAGLPGAK-----GLTSPSPGPPGDKTGPDPGAGQDGRP 300
Db 754 GKDGVRGLTGTPGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 813
QY 301 GPPPPGARGQ-----AGVMGFPCKAAGPGKAGRGVPGPPGA----- 341
Db 814 GADGCPGAKGPGDAGAKGDAGPPGAPGPPGICNVGAPGAKGARGAGPPGATGFP 873
QY 342 -----VGPAGKDGAGAGQGGPPGAPG---AGERGEQPGAPSPGQGLPGPAGPPGAGK 393
Db 874 GAAGRVGPPGSPGNAGPPGPPGAGKGGKRGRTGPGAGPGEVGGPPGPPGAGKGGSP 933
QY 394 GEQGVPGDLGAPGSPGAPGPPGTGLPGPGERGPGSRGPPGADGVAGPKGAPAGRGSP 453
Db 938 GADGPAGAGTGGPQGIAGORGVVGLPGQGRGPPGLPGPSPGPPGKQSPGSGASGPP 997
QY 454 GPAGP---KSGPGEAGRPGAGLPGAKGLTGSPGSPGPDGKTGTPPGAGQDGRPPGPP 510
Db 998 GPMGPPGLAGPPGSGREGAPGARGSPGDRGSPGAKGDRGTGPGAGPAGPAGPAGPV 1057
QY 511 GARGQAGVMGPPGKGAAGPGKAGRGVPGPPGAVGAPGKDGPE-----AGAQ 558
Db 1058 GPAGKSGDRGTGPGAGPAGPVGARGPAGPQGRGDKGTGEGQDRGKIGHRGFSLG 1117
QY 559 GPPGPP-----AGPAGERGEGPAGSP---GFQGLPGPAGPPGAGKGGVGP 603
Db 1118 GPPGPPGSPGQSPGSGAGPPGPPGSPGAPGKQGLNGLPGPPGPPGRTGDRGPV 1177
QY 604 GDLGAPGSPGAP 616
Db 1178 GPPGPPGPPGPPG 1190


```

QY 58 GRPGEAGLP---GAKGLTSGSPSGP-----DGKTPPGPAGQDGRPPGPPGAR 105
DB 239 GRPGERPPGQAGRLPGTAGLPGMKHGRFSGLDGAKDAGPAGPKGPFSGENGAP 298
QY 106 QGAGVMGFPFGKGAAGPBGKAGER-----GVPPPGAVGPA-----GKDGAGAG 150
DB 299 QMGPRGLPGRGPPGAPGAPAGKNGDGTAGTAAAGPPGTGPAGPPGPGAVGAKGAPQ 358
QY 151 GPPGAPGAPGRGCGP-----AGSPFGGLPGPA 180
DB 359 GPRGSEPGQVGRGFPFGPPGAGAGPAGNPGADQCGAKGANGAPGAGPAGPFGARGPS 418
QY 181 GPPGAGKPGQGVPGDLGAP-----GSPGPA-----GEPGPT 213
DB 419 GPQPGGPPGPKNSGSEPGAGCSKGDTCAGKEGPPGVQGGPAGGEGKRGARGEPPT 478
QY 214 GLPDPGGRGPGSRGPPGADGVAGPKPAGERGSPGAPKSGPAGSPGAGPGEAGLPGAK 273
DB 479 GLPDPGGRGPGSRGPPGADGVAGPKPAGERGSPGAPKSGPAGSPGAGPGEAGLPGAK 538
QY 274 GLTSPGSGPDGKTPGPPAGQDGRPPGPPGARGQAGVMGPPGKGAAGEPKAGER 333
DB 539 GLTSPGSGPDGKTPGPPAGQDGRPPGPPGARGQAGVMGPPGKGAAGEPKAGER 598
QY 334 GVPDPGAVGAPAGKDGAGAGQPPGPPGAPGAGERGQPPGPGFQGLPGPAGPGEAGKP 393
DB 599 GVPDPGAVGAPAGKDGAGAGQPPGPPGAPGAGERGQPPGPGFQGLPGPAGPGEAGKP 658
QY 394 GEQVPGDLGAPSGPAGE-----PGP-----TGLPGPP 423
DB 659 GEQVPGDLGAPSGPAGE-----PGP-----TGLPGPP 423
QY 424 GERGPGSRGPPGADGVAGPKPAGERGSPGAPKSGPAG-----AGRPGEAGLP 474
DB 719 GSQAPGLQMGPPGEGAAGLFGPKDRDADGFKADSGKDGVRGLTGTGPPGAPAG 778
QY 475 GAK-----GLTSPGSGPDGKTPGPPAGQDGRPPGPPGARGQ----- 515
DB 779 GDKGESGSPGAPGTGARGAPGDRGEPGPPGAPGPPGADGQPGAKGEPGDAGAKGDA 838
QY 516 -----AGVMGPPGKGAAGECKAGERGVPPGGA-----VGPAGKDGAGAGOPP 561
DB 839 GPPGAPGAPGPPGTGNVGAPEAKGARGSPGATGPPGAAGRVGPPGSGNAGPPGPP 898
QY 562 GPAGS---AGRGEQGPAGSPGFGQLPGPAGPGEAGKPGGQGVPGDLGAPGSPGAG 616
DB 899 GPAGKEGKGPRGTGPRGPRGVEVPPGPPGPPGAGEKSGADGAPAGPTGPGQGIAG 956

RESULT 7
Q63079
ID Q63079 PRELIMINARY; PRT; 1453 AA.
AC Q63079;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Collagen alpha1 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Bone, and Tooth;
RX MEDLINE=99163824; PubMed=10065941;
RA Brandsten C., Lundmark C., Christersson C., Hammarstrom L., Wurtz T.;
RT "Expression of collagen alpha1(I) mRNA variants during tooth and bone
RT formation in the rat";
RL J. Dent. Res. 78:11-19(1999).
DR EMBL; Z78279; CAB01633.1; -
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.

```

```

DR GO; GO:0005817; P:phosphate transport; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; Fib_collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD0000007; C1g_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VMC; 1.
DR PROSITE; PS01208; VWF_1; 1.
DR PROSITE; PS0184; VWF_2; 1.
KW Collagen.
FT NON_TER
SQ SEQUENCE 1453 AA; 137886 MW; E6896BDC19A4A1D8 CRC64;

Query Match 63.5%; Score 2215.5; DB 2; Length 1453;
Best Local Similarity 53.8%; Pred. No. 7.2e-82;
Matches 446; Conservative 21; Mismatches 149; Indels 213; Gaps 16;

QY 1 GPPGE---PGPTGLPGPPGGRGSGRFPF-----ADGVAGPKPAGER 42
DB 117 GPPGQDGIPIPGPLPGPPGPPGPPGPPGLGNGNFASQMSYGYDEKAGAGVSPGMGPPGPR 176
QY 43 GSPGAPAGKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTPGPPGAGQDGRP---GPP 99
DB 177 GLPDPGARGPQGFQGPPEGPEGSGSPMGPPGPPGPKNGDDGEGACKRPPGREGPP 236
QY 100 GPPGARGQGVMGFP-----GPKGAAGEPGK-----A 126
DB 237 GPQARGLPGTAGLPGMKHGRFSGLDGAKDGTGTPAGPKGPGSPGSENGTTPQMGPRGLP 296
QY 127 GERGVPPGAVGAPAGKDGAGAGQPPG-----AGPAGERGQGP- 167
DB 297 GERGRPPPTAGARGNDGAVGAAGPPGTGTGTPGPPGGAAGKAGEAGPQARGSEGPQ 356
QY 168 -----AGSPGFQGLPGPAGPGEAGKP 189
DB 357 GVRGEPGPPGAPAGAPNPGADGPGAKGANGAPGAGPFGPAGPSPGQPSGAP 416
QY 190 GEGVPGDLGAP-----GSPGPA-----GEPGTLGLPGPPGR 222
DB 417 GPKGTSGEPGAPGNKGDTCAGKEPAGVQGGPAGGEGKRGARGEPGSLGPPGPR 476
QY 223 GPGSGRFPGADGVAGPKPAGERGSPGAPKSGPGEAGRPGEAGLPGAKGLTSGSPGP 282
DB 477 GPGSGRFPGADGVAGPKPAGERGSPGAPKSGPGEAGRPGEAGLPGAKGLTSGSPGP 536
QY 283 GPDGKTPGPPGAPQDGRPPGPPGARGQAGVMGPPGKGAAGEPKGAGERGVPPGPAV 342
DB 537 GPDGKTPGPPGAPQDGRPPGPPGARGQAGVMGPPGKGAAGEPKGAGERGVPPGPAV 596
QY 343 GPAGKDGAGAGQPPGPPGAPGAGERGQGPAGSPGFGQLPGPAGPGEAGKPGGQGVPGDL 402
DB 597 GPAGKDGAGAGQAPGAPGAPGAGERGQGPAGSPGFGQLPGPAGPGEAGKPGGQGVPGDL 656
QY 403 GAPCPSPGAGE-----PGP-----TGLPDPGGRGPPGPR 432
DB 657 GAPCPGARGRFPGERGVQGGPAGPRGNNGAPGNDCAKGDGTGAPGAPSGQAGPLQ 716
QY 433 GPGCAGDVAGPKPAGERGSPGAPKSGPGE-----AGRPGEAGLPGAK----- 477
DB 717 GMPGEGAAGLPGPKGDRDAGPKGADGSPGKGVGRGLTGTPIGPPGAPAGDKGAGSPS 776
QY 478 ---GLTSGSPGPDGKTPGPPGAGQDGRPPGPPGARGQ-----AGVM 519
DB 777 GPAGPTGARGAPGDRGEAGPFPAGPAGPPGADGQPGAKGEPGDTGVKGDAGPPGAPGA 836
QY 520 GPGPKGAAGEPKKAGERGVPPGPA-----VGPAGKDGAGAGQPPGAPG---A 567
DB 837 GPGPTGNVGAAPGPKGSRGAAGPPGATGTPGGAAGRVGPPGPGSNAGPPGPPGVKGGK 896

```


Db	1047	GPAGKNGDRGETGPAGPAGTGPAGARGPAGPQGRGDKGTGQGGDRGIKGRHGFSLQ	1106
Qy	547	GPAGKDEACAGCPGPPGAPAGERGEQGPAGSP--CFQGLPGPAGPPGAGKPGEGQVP	603
Db	1107	GPGSPSPSGEQPFGSAGSPAGKPGPFGSAGSPGKGLNLPFGTIGPPGPRGRGTGDSGPA	1166
Qy	604	GDLAGCPSPGPAG	616
Db	1167	GPPGPPGPPGPPG	1179

RESULT 10			
6PCL3			
ID	G6PCL3	PRELIMINARY;	PRT; 1225 AA.
AC	G6PCL3;		
DT	05-JUL-2004	(TrEMBLrel. 27, Created)	
DT	05-JUL-2004	(TrEMBLrel. 27, Last sequence update)	
DT	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)	
DE	Collal protein.		
GN	Name=Collal;		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
[1]			
RN	SEQUENCE FROM N.A.		
RP	STRAIN=Mix FVB/N;		
RC	TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;		
RC	MEDLINE=22388257; PubMed=12477932;		
RA	Stearnsberg R.L., Feingold E.A., Grouse L.H., Derse J.G.		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Hopkins S.F., Zeeberg B., Bucow K.H., Schaefer C.F., Bhat N.K.,		
RA	Altschul R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,		
RA	Brownstein L.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Whiting M., Helton E., Kettelman M., Madan A.C., Rodrigues S., Sanchez A.,		
RA	Blakesley R.W., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Rodriguez A.C., Trinchman J.W., Green E.D., Dickson M.C.,		
RA	Krzywinski M.I., Skalska U., Smalus D.E., Scherch A., Schein J.E.,		
RA	Jones S.J., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RL	and mouse cDNA sequences."		
RT	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		

RP	SEQUENCE FROM N.A.
RC	STRAINMix FVB/N;
RC	TISSUE= Mammary tumor. WAP-TGF alpha model. 7 months old;
RR	Strasbourg R.;
RR	Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RR	EMBL; BC059281; AAHS9281.1; -
DR	InterPro; IPRO0161; Cig.helix.
DR	InterPro; IPRO0160; Collagen.
DR	InterPro; IPRO02181; Fibcollagen_C.
DR	InterPro; IPRO00885; Fib.collagen_C.
DR	InterPro; IPRO03041; PMP_SGCI.
DR	InterPro; IPRO01007; VWF_C.
DR	Pfam; PF01410; COLFI; 1.
DR	Pfam; PF01391; Collagen; 14.
DR	Pfam; PF00093; WVC; 1.
DR	ProDom; PDC000007; Cig.helix; 2.
DR	ProDom; PD002078; Fib.collagen_C; 1.
DR	SMART; SM00038; COLFI; 1.
DR	SMART; SM00214; WVC; 1.
DR	PROSITE; PS01208; VWF_C.1; 1.
DR	PROSITE; PS0184; VWF_C.2; 1.
KW	Collagen.

```

SQ SEQUENCE 1225 AA; 117860 MW; B6386CBB4457F4D9 CRC64;

Query Match 62.4%; Score 2176.5; DB 2; Length 1225;
Best Local Similarity 56.7%; Pred. No. 2.4e-80;
Matches 431; Conservative 28; Mismatches 157; Indels 144; Gaps 133

QY 1 GPPGEGPTGLPPGE-----RGPGRFPFGADGVAGFKGPAGRGSPGA 48
DB 180 GPPGAPPGQFGPPGEPGEGSGSPGPRPPGPKNGDDGEAGKPGRGERGGPGQ 239
QY 49 GPKGSGPAGRP-----GEAGLFGAKGLTGSPPSPGPDGKLTGPPGAGOD 93
DB 240 GARGLPFGTAGLPGMKHGRFGSLGDAKGDAPGAPKGPSPGNGAPQMGRLPG 299
QY 94 GRPGPPGPPGARGQAGVNGFPKGAAGE-----PGKAGRGVPPGPAVGPAGK 147
DB 300 GRPGPPGTAGRGNDGAVGANGPPGTCTPCPFCAVCAKGEAGPQAGRSEGP 359
QY 148 GAQGPFGPAGPAGERGEGP-----AGSPGQFLPGPAGPPGPAEAKPG 192
DB 360 GEPGPGPAGAAGPAGNPGADQCPGAKGANGAPGIAGAPFGPARGSPGQSP 419
QY 193 GVPGLDGLAP-----GPSGA-----CEPPTGLPDPGPPBERGP 225
DB 420 GNSGFGPAPGNKDTCAKGEPTGCVGPPGPPAGEBGRGARPEPFGLP 479
QY 226 GSRGFPAGDVAGPKPAGERGSPGAPGKSPGEEAGRPGEAGLPGAKGLTGS 285
DB 480 GSRGFPAGDVAGPKGSPGERGADPGAPGKSPGEEAGRPGEAGLPGAKGLT 539
QY 286 GXTGPPGAGODGRPPGPPGARGQAGVNGFPKGTAGEPKAGRGVPPGPA 345
DB 540 GXTGPPGAGODGRPPGPPGARGQAGVNGFPKGTAGEPKAGRGVPPGPA 599
QY 346 GKDGEAGAQGPPGAPAGERGEGQPSGSPGQGLPGPAGPPGEEAGKPG 405
DB 600 KDGGEAGAQGAPGAPAGERGEGQPSGSPGQGLPGPAGPPGEEAGKPG 659
QY 406 GPSGAGE-----PGP-----TCLPGPBERGGRPSRGFP 435
DB 660 GPSGARGBRGFPGRGVQGPFGPPGPRGNNGAPENDCAKGDGTGAP 719
QY 436 GADGVAGPKGPAGBRGSPGAPKGSPE-----AGRPGEAGLPGAKGLT 486
DB 720 GERGAAGLPGPKGRDAGPKGADGSPGKDARGLTGPIGPPGAPAGCD 779
QY 487 GPDGKTGPPAGODGRPPGPPGARGQAGVNGFPDPKGAAGEPKAGRGV 546
DB 780 GPTGARGAFGRGAGPPGAPGAPGAPGAFGAPGVPGAPKNGDRGET 839
QY 547 GPGAKDGBAGAQGPPGAPAGERGE-----QGPAKSPGQSLPGPA----- 588
DB 840 GPAGARGAPGQGRGDKGTGEGQDGRGKGRHGRFSGLGQPPGSPG 899
QY 589 ---GPPBAGKPE---QGVPPDLGAPG-----SGPAG 616
DB 900 GPRGPPSAGSPGKDLGLNGLPPIGPPGPRGRTGDSGAP 939

```

RESULT	11
AAH59281	PRELIMINARY; PRT; 1225 AA.
ID	AAH59281
AC	AAH59281;
DT	02-MAR-2004 (TrEMBLrel. 27, Created)
DT	02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT	02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE	Collal protein.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
[1]	..
RN	SEQUENCE FROM N.A.
RP	

DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 4.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C.1; UNKNOWN_1.
DR PROSITE; PS0184; VWF_C.2; 1.
DR Collagen.
KW COLLAGEN.
SQ SEQUENCE 1450 AA; 137563 MW; ABP8A74841B87B7C CRC64;

Query Match 58.6%; Score 2043.5; DB 2; Length 1450;
Best Local Similarity 54.7%; Pred. No. 5.7e-75;
Matches 401; Conservative 31; Mismatches 184; Indels 117; Gaps 11;

QY 1 GPP-----GEFGTGLPDPGERGSGRGGPFGADGVAGPKGAGERGSGPA 48
DB 444 GPPGSGEGKRGSRGEPGAPGPPFAGGERGSGRGGPFGSDGASGPKGAGERGSGVGA 503
QY 49 GPKSGPAGRGEAGLPCAKGLTSGSGSPGPDGKTGPPGAGQDGRGPPGPPGARGQA 108
DB 504 GPKGTSGRGEPGLPCAKGLTSGSGSPGPDGKTGAGAGQDGHGPPGSGARGQS 563
QY 109 GVMGPPGKGAAGEFGKAGERGVPPGAVGAGKDGAGACGPPGPPGAGERGEGQPA 168
DB 564 GVMGPPGKGAAGEFGKAGERGVPPGATGAPKDGAGACGPPGPPGSGRGERGQPA 623
QY 169 GSPGQGLGPPAGPGEAGKAGEQGVGDGLGAPGSGPAGE-----PGTGLPDP- 218
DB 624 GSPGQGLGPPAGPGEAGKAGEQGVGDGLGAPGSGPAGE-----PGTGLPDP- 218
QY 219 -----PGERGSGRGGP----- 231
DB 684 GSPGSGNDGAKGEAGAGACGGRGPPGLQGPPGERGSGAGMPGAKGDRGDAGTKGADGAP 743
QY 232 ----GADGVAGPKGAGERGSP-----GAPGKSGPGEAGRGEAGLPCAKGLTSP 279
DB 744 GKDGAAGLTGTPGPPGSGAPQDKGEGGSPGAGTARGSGFERGEPGAPGAGTICGPP 803
QY 280 GS---PGPDGKTGPPGAGQDGRGPPGPPGARGQAGVMGPPGKGAAGEFGKAGERGV 336
DB 804 GADGQKAGEGSDAGPKDAGAPGAPGTGAPGAGNVGAPGKTRGAAGPPGATGFP 863
QY 337 GPPGAVGPKDGEAGAGAGGPPGAPGAGERGQGGPAGSGPFOGLGPPAGPP---GEAGKP 393
DB 864 GAAGRLGPPGSGNAGPPGPPGKGEAKGSRGEGTGPAGRSSEPPGAPGPPSGERKSP 923
QY 394 GEGVPGDLGAPGSPGAPGPTGLPGPPGERGSGRGGPFGADGVAGPKGAGERGSP 453
DB 924 GSDGPAGAPGIPGPGIACQGVVGLPGQGERGSGGLPGPAGEPKQGPSPGNGERGPP 983
QY 454 GPAGP-----KSGPGEAGRPGAGLPGAK-----GLTSGSPGSPGDKTGP 495
DB 984 GFGSGPGLGGPPGEGPREGSPGSEGAPGRDGSFGPKGDRGNGSPGPPGAPGAPGAPV 1043
QY 496 GPAGQD---GRPPGPPGARGQAGVMGPPGKGAAGEFGKAGE-----RGVPGPP 543
DB 1044 GPAGKNGDRGEGTGPAGPAGPSGVRGAPGAGARGDRGAGEQGERGMKGRHGFNGW 1103
QY 544 GAVGPAGKDGAGAGAGQPPGAPGAGERGQGGPFGQGLPGPAGPPGEGAGKPGEGQVP 603
DB 1104 GPPGPPGSGGEGQAGPSPGAPGPPGSGSGTGDKGVNGLPGIPGPPGRCRNGDVGA 1163
QY 604 GDLGAPGSGPAG 616
DB 1164 GPPGPPGPPGPPG 1176

RESULT 14
O93251

ID O93251 PRELIMINARY; PRT; 1445 AA.
AC O93251;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen.
GN Name=alpha 1 type I collagen;
OS Rana catesbeiana (Bull frog);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranidae; Rana.
OX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95294154; PubMed=10367734;
RA Asahina K., Utoh R., Obara M., Yoshizato K.;
RT "Cell-type specific and thyroid hormone-dependent expression of genes
of alpha 1(I) and alpha 2(I) collagen in intestine during
amphibian metamorphosis";
RL Matrix Biol. 18:89-103(1999).
DR EMBL; AB015440; BAA29028.1; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 4.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C.1; 1.
DR PROSITE; PS0184; VWF_C.2; 1.
KW Collagen.
SQ SEQUENCE 1445 AA; 137251 MW; P59B5550C9873F04 CRC64;

Query Match 58.6%; Score 2042.5; DB 2; Length 1445;
Best Local Similarity 49.7%; Pred. No. 6.2e-75;
Matches 421; Conservative 34; Mismatches 161; Indels 231; Gaps 18;

QY 1 GPPGEPGFTGLPDPGERGSGS-----RGFFGADGVAGPKGPA-----GER 42
DB 221 GKPRGPRGPPGPPGQARGLPGTAGLPGMKGRHGFNGLDGAKGDTGPAGPKGPPNGEN 280
QY 43 GSPGAPGKSGPGEAGRPGAGLPGAKLTGSGSPGPDGKTGPPG-----A 90
DB 281 GAGQGVPRGLPGERGPPSGPAGARGNDGTPGAAGPPGPTGTPPPGPGVGPAGDA 340
QY 91 GQDGRPPPPGARGQAGVMGPPGKGAAGEFGKAGERGVPPGAVGPKDGEAGQA 150
DB 341 GPQSRGPPGPPGARGPPGAGPPGAGPAGSAGNFGTDGQPCAKGATGAPGIAGAPPPCAR 400
QY 151 GPPGAPGAGERGQGGPAGSGPGLPGPAGPPGEGACKPQEQVPGDLGAPGSGPAGEP 210
DB 401 GAGPQGGSGPPGKGNNGEPGAGKGEKAGEGSPAGSGPPGPPGEGKRGSRGEP 460
QY 211 GPTGLPDPGERGSGRGGPFGADGVAGPKGAGERGSPGAPGKSGPGEAGRPGEGALP 270
DB 461 GPSPPGAPAGERGAPGSRGPPGADGACGPKGPPGERGPPVGSAGFKSGSGRPPGLP 520
QY 271 GAKGLTSGSGPDGKTGTPPGAGQDGRGPPGPPGARGQAGVMGPPGPKGAAGEFGKA 330
DB 521 GAKGLTSGSGPDGKTGTPAGAGQDGRGPPGPPGARGQAGVMGPPGPKGAAGEFGKP 580
QY 331 GERGVPPGAVGPKDGEAGAGQGGPPGAGERGEGQGPAGSGPFGQLGPPGAPGPEGA 390
DB 581 GERGVPPGAVGPKDGEAGAGQGGPPGAGERGEGQGPAGSGPFGQLGPPGAPGPEGA 640
QY 391 GKPGEQGVPGD-----LGAPGSGP-----AGEP 414

Tue Oct 19 09:27:59 2004

Qy	541	GPPGAVGPAGKGEAGAGQGPFPAGPA-----GERGEQFPAGSFPQ	582
Db	1038	GAPGVGPAGKSGDRGETGSPGAPAGTAGARGPAGPQGPGRGDKGEAGEQGERGMKGR	1097
Qy	583	GLPGPAGPGEAGKPGEGVPGDLGAPGPSGPAG	616
Db	1098	GFNGSPGPFPFGSSGEQFSGASGPAGPRGPPG	1131

Search completed: October 18, 2004, 13:41:50
Job time : 117.271 secs

This Page Blank (uspto)

CC pharmaceutically active compounds. After administration, the coupled
 CC medicament will not diffuse from the circulating blood into the
 CC interstitium. Clearance by liver and kidney will be kept to a minimum,
 CC ensuring a more constant plasma level of the medicament. Suitable
 CC medicaments include those involved in intervening blood clotting,
 CC vasodilation, function of erythrocytes, thrombocytes and leukocytes,
 CC immune responses, and blood levels of messenger molecules such as
 CC hormones.
 XX
 SQ Sequence 821 AA;
 Query Match 100.0%; Score 4640; DB 8; Length 821;
 Best Local Similarity 100.0%; Pred. No. 3e-278;
 Matches 821; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GPPGEPGPTGLGPPGRCGPGSRGFPAGADGAGKGPAGRGSPGAPGKSGPCEAGRP 60
 Db 1 GPPGEPGPTGLGPPGRCGPGSRGFPAGADGAGKGPAGRGSPGAPGKSGPCEAGRP 60
 QY 61 GEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGAGQAGVMGFPKGA 120
 Db 61 GEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGAGQAGVMGFPKGA 120
 QY 121 GPPGKAGRGVPPGAVGPPGAGKGEAGAGQAGPPGPPGAGRGSGPAGSGPGLPGPA 180
 Db 121 GPPGKAGRGVPPGAVGPPGAGKGEAGAGQAGPPGPPGAGRGSGPAGSGPGLPGPA 180
 QY 181 GPPGKAGRGVPPGAVGPPGAGKGEAGAGQAGPPGPPGAGRGSGPAGSGPGLPGPA 240
 Db 181 GPPGKAGRGVPPGAVGPPGAGKGEAGAGQAGPPGPPGAGRGSGPAGSGPGLPGPA 240
 QY 241 GPAGRGSPGAPGKSGPCEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRP 300
 Db 241 GPAGRGSPGAPGKSGPCEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRP 300
 QY 301 GPPGPPGARGAGVPPGKGAAGEPKKAGRGVPPGPPGAGVAGKDGAGAGCGPPGPA 360
 Db 301 GPPGPPGARGAGVPPGKGAAGEPKKAGRGVPPGPPGAGVAGKDGAGAGCGPPGPA 360
 QY 361 GPAGRGSGPAGSGPGLPGPAGPPGAGKGEAGQVPPGDLGAPGSPGAPGPPGPTGLP 420
 Db 361 GPAGRGSGPAGSGPGLPGPAGPPGAGKGEAGQVPPGDLGAPGSPGAPGPPGPTGLP 420
 QY 421 GPPGRCGSPRGPPGADGAGKGPAGRGSPGAPGKSGPGEAGRPBAGLPGAKGLT 480
 Db 421 GPPGRCGSPRGPPGADGAGKGPAGRGSPGAPGKSGPGEAGRPBAGLPGAKGLT 480
 QY 481 GSPGSPGPDGKTGPPGAGQDGRPPGPPGARGAGQAGVMGFPKGAAGPPKAGRGVP 540
 Db 481 GSPGSPGPDGKTGPPGAGQDGRPPGPPGARGAGQAGVMGFPKGAAGPPKAGRGVP 540
 QY 541 GPPGAVGPPAGKDGAGAGQAGPPGPPGAGRGSGPAGSGPGLPGPAGPPGAGKPEQ 600
 Db 541 GPPGAVGPPAGKDGAGAGQAGPPGPPGAGRGSGPAGSGPGLPGPAGPPGAGKPEQ 600
 QY 601 GVPDGLGAPGSPGAPGPPGPTGLPGRGSGRFPAGADGAGKGPAGRGSPGPA 660
 Db 601 GVPDGLGAPGSPGAPGPPGPTGLPGRGSGRFPAGADGAGKGPAGRGSPGPA 660
 QY 661 GPKGSPGAGRPAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGAGQ 720
 Db 661 GPKGSPGAGRPAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGAGQ 720
 QY 721 GVMGFPKGAAGPPKAGRGVPPGAVGPPGAGKGEAGAGQAGPPGPPGAGRGSGPA 780
 Db 721 GVMGFPKGAAGPPKAGRGVPPGAVGPPGAGKGEAGAGQAGPPGPPGAGRGSGPA 780
 QY 781 GSPGFGQLPGPAGPPGAGKPEQGVGPDGLGAPGSPGAGG 821
 Db 781 GSPGFGQLPGPAGPPGAGKPEQGVGPDGLGAPGSPGAGG 821

ADMA48391
 ID ADM48391 standard; protein; 617 AA.
 XX AC ADM48391;
 XX DT 03-JUN-2004 (first entry)
 XX DE Human recombinant gelatin-like polypeptide Hu-3.
 XX KW Plasma substitute; Gelatin-like protein; plasma expander; human.
 XX OS Homo sapiens.
 XX PN EP1398324-A1.
 XX PD 17-MAR-2004.
 XX PF 11-SEP-2002; 2002EP-00078745.
 XX PR 11-SEP-2002; 2002EP-00078745.
 XX PA (PUUF) FUJI PHOTO FILM BV.
 XX PI Bouwstra JB, Toda Y;
 XX WP 2004-229415/22.
 XX PS Example 1; SEQ ID NO 2; 31pp; English.
 XX CC The present sequence is the protein sequence of recombinant gelatin-like
 CC protein Hu-3. This is a trimer of human recombinant gelatin-like protein
 CC Hu-1 ADM48390. A claimed composition suitable as a substitute for plasma
 CC comprises a solution of saline and a protein having a colloid osmotic
 CC function. The protein is a recombinant gelatin-like protein with a
 CC molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric
 CC point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The
 CC recombinant gelatin-like protein, or a dimer, trimer or tetramer of the
 CC protein, is useful as a plasma expander that has a lower clearance rate
 CC from blood circulation, provides better and predictable regulation of
 CC clearance rate and which is less susceptible to proteolytic degradation
 CC than presently used gelatin derivatives. Recombinant gelatin-like
 CC proteins that are in essence free of hydroxyproline do not give rise to
 CC an immunological reaction with blood samples containing IgG antibodies.
 CC The gelatin-like proteins can be covalently attached to pharmaceutically
 CC active compounds. After administration, the coupled medicament will not
 CC diffuse from the circulating blood into the interstitium. Clearance by
 CC liver and kidney will be kept to a minimum, ensuring a more constant
 CC plasma level of the medicament. Suitable medicaments include those
 CC involved in intervening blood clotting, vasodilation, function of
 CC erythrocytes, thrombocytes and leukocytes, immune responses, and blood
 CC levels of messenger molecules such as hormones.
 XX
 SQ Sequence 617 AA;
 Query Match 75.0%; Score 3482; DB 8; Length 617;
 Best Local Similarity 100.0%; Pred. No. 5.1e-207;
 Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GPPGEPGPTGLGPPGRCGPGSRGFPAGADGAGKGPAGRGSPGAPGKSGPCEAGRP 60
 Db 1 GPPGEPGPTGLGPPGRCGPGSRGFPAGADGAGKGPAGRGSPGAPGKSGPCEAGRP 60
 QY 61 GEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGAGQAGVMGFPKGA 120
 Db 61 GEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGAGQAGVMGFPKGA 120
 QY 121 GPPGKAGRGVPPGAVGPPGAGKGEAGAGQAGPPGPPGAGRGSGPAGSGPGLPGPA 180
 Db 121 GPPGKAGRGVPPGAVGPPGAGKGEAGAGQAGPPGPPGAGRGSGPAGSGPGLPGPA 180

QY 181 GPPGEAGKPGEGVQVVDLAPSGSPAGEPGTGLPGPPGGERGGPSRGGPPGADGVAGPK 240
DB 181 GPPGEAGKPGEGVQVVDLAPSGSPAGEPGTGLPGPPGGERGGPSRGGPPGADGVAGPK 240
QY 241 GPAGERSGPPAGPKGSPGAGRPGEAGLPGAKGLTSGPSGPDGKTGTPPGAGQDGRP 300
DB 241 GPAGERSGPPAGPKGSPGAGRPGEAGLPGAKGLTSGPSGPDGKTGTPPGAGQDGRP 300
QY 301 GPPGPPGARGQAGVWMPFGPKGAAGEPGKAGRGVPPGPAVGPAGKDGAGAGQPPGPA 360
DB 301 GPPGPPGARGQAGVWMPFGPKGAAGEPGKAGRGVPPGPAVGPAGKDGAGAGQPPGPA 360
QY 361 GPAGERGEQAGSPGPGQGLPGPAGPGEAGKPGEGVQVVDLAPSGSPAGEPGTGLP 420
DB 361 GPAGERGEQAGSPGPGQGLPGPAGPGEAGKPGEGVQVVDLAPSGSPAGEPGTGLP 420
QY 421 GPPGERGGPSRGGPPGADGVAGPKGSPAGEPGTGLPGPPGGERGGPSRGGPPGADGVAGPK 480
DB 421 GPPGERGGPSRGGPPGADGVAGPKGSPAGEPGTGLPGPPGGERGGPSRGGPPGADGVAGPK 480
QY 481 GSPGSPGPDGKTGTPPGAGQDGRPDPGPPGARGQAGVWMPFGPKGAAGEPGKAGRGV 540
DB 481 GSPGSPGPDGKTGTPPGAGQDGRPDPGPPGARGQAGVWMPFGPKGAAGEPGKAGRGV 540
QY 541 GPPGAVGPAGKDGAGAGQPPGPAVGPAGERGEQAGSPGPGQGLPGPAGPGEAGKPGEG 600
DB 541 GPPGAVGPAGKDGAGAGQPPGPAVGPAGERGEQAGSPGPGQGLPGPAGPGEAGKPGEG 600
QY 601 GVPDGLGAPGSPGAG 616
DB 601 GVPDGLGAPGSPGAG 616

RESULT 3
ADQ26217 standard; protein; 617 AA.
XX ADQ26217;
XX AC ADQ26217;
XX DT 23-SEP-2004 (first entry)
XX DE Human gelatine-like polypeptide Hu-3.
XX KW Human; gelatine-like protein; Hu-3; microcarrier; cell culture.
XX OS Homo sapiens.
XX PN WO2004056976-A2.
XX PD 08-JUL-2004.
XX PF 23-DEC-2003; 2003WO-NL000922.
XX PR 23-DEC-2002; 2002EP-00080539.
XX PA (FUUF) FUJI PHOTO FILM BV.
XX PI Bouwstra JB, Van Es AJJ, Toda Y;
XX PS WPI; 2004-507711/48.
XX PT Preparing cell culture support useful for culturing anchorage dependent
PT cells, involves coating microcarrier bead with gelatine or gelatine-like
PT protein having specific molecular weight.
XX Example 1; SEQ ID NO 1; 19pp; English.

CC The present sequence is the protein sequence of human recombinant gelatin
CC -like polypeptide Hu-3, which has a molecular weight of approximately 54
CC kDa and which contains multiple Gly-Xaa-Yaa triplets. In an example from
CC the invention, recombinant Hu-3 was immobilised on polystyrene beads
CC using a heterobifunctional crosslinking agent, and used in a spinner
CC flask cell culture. A claimed process for the preparation of a cell

CC culture support comprises the step of coating a microcarrier bead with
CC gelatine or gelatine-like protein having a molecular weight of about 40-
CC 200 kDa, and optionally further comprising the step of immobilising the
CC gelatine or gelatine-like protein on the microcarrier. In this process,
CC more than 75%, preferably more than 85% and more preferably more than 95%
CC of the gelatine or gelatine-like protein has the same molecular weight.
CC The gelatine or gelatine-like protein is recombinantly produced to obtain
CC a material of uniform molecular weight and to reduce the risk of
CC contamination with prions. A claimed cell support consists of microbeads
CC of 50-500 um size coated with a gelatine-like protein consisting of at
CC least 95% Gly-Xaa-Yaa triplets and containing at least 15% proline
CC residues and less than 5% of hydroxyproline residues, with a molecular
CC weight distribution showing a maximum between 40 and 200 kDa, at least
CC 75% of the protein molecules having a molecular weight within 2% of the
CC maximum. Large-scale production of expressed products can be accomplished
CC with gelatine-coated microcarriers.

XX Sequence 617 AA;
SQ
Query Match 75.0%; Score 3482; DB 8; Length 617;
Best Local Similarity 100.0%; Pred. No. 5.1e-207;
Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GPPGPPGPTGLPGPPGGERGGPSRGGPPGADGVAGPKGSPAGEPGTGLPGPPGGERGGPSRGGPPGADGVAGPK 60
DB 1 GPPGPPGPTGLPGPPGGERGGPSRGGPPGADGVAGPKGSPAGEPGTGLPGPPGGERGGPSRGGPPGADGVAGPK 60
QY 61 GEAGLPGAKGLTSGPSGPDGKTGTPPGAGQDGRPDPGPPGARGQAGVWMPFGPKGAA 120
DB 61 GEAGLPGAKGLTSGPSGPDGKTGTPPGAGQDGRPDPGPPGARGQAGVWMPFGPKGAA 120
QY 121 GPPGKAGRGVPPGPAVGPAGKDGAGAGQPPGPAVGPAGERGEQAGSPGPGQGLPGPAG 180
DB 121 GPPGKAGRGVPPGPAVGPAGKDGAGAGQPPGPAVGPAGERGEQAGSPGPGQGLPGPAG 180
QY 181 GPPGEAGKPGEGVQVVDLAPSGSPAGEPGTGLPGPPGGERGGPSRGGPPGADGVAGPK 240
DB 181 GPPGEAGKPGEGVQVVDLAPSGSPAGEPGTGLPGPPGGERGGPSRGGPPGADGVAGPK 240
QY 241 GPAGERSGPPAGPKGSPGAGRPGEAGLPGAKGLTSGPSGPDGKTGTPPGAGQDGRP 300
DB 241 GPAGERSGPPAGPKGSPGAGRPGEAGLPGAKGLTSGPSGPDGKTGTPPGAGQDGRP 300
QY 301 GPPGPPGARGQAGVWMPFGPKGAAGEPGKAGRGVPPGPAVGPAGKDGAGAGQPPGPA 360
DB 301 GPPGPPGARGQAGVWMPFGPKGAAGEPGKAGRGVPPGPAVGPAGKDGAGAGQPPGPA 360
QY 361 GPAGERGEQAGSPGPGQGLPGPAGPGEAGKPGEGVQVVDLAPSGSPAGEPGTGLP 420
DB 361 GPAGERGEQAGSPGPGQGLPGPAGPGEAGKPGEGVQVVDLAPSGSPAGEPGTGLP 420
QY 421 GPPGERGGPSRGGPPGADGVAGPKGSPAGEPGTGLPGPPGGERGGPSRGGPPGADGVAGPK 480
DB 421 GPPGERGGPSRGGPPGADGVAGPKGSPAGEPGTGLPGPPGGERGGPSRGGPPGADGVAGPK 480
QY 481 GSPGSPGPDGKTGTPPGAGQDGRPDPGPPGARGQAGVWMPFGPKGAAGEPGKAGRGV 540
DB 481 GSPGSPGPDGKTGTPPGAGQDGRPDPGPPGARGQAGVWMPFGPKGAAGEPGKAGRGV 540
QY 541 GPPGAVGPAGKDGAGAGQPPGPAVGPAGERGEQAGSPGPGQGLPGPAGPGEAGKPGEG 600
DB 541 GPPGAVGPAGKDGAGAGQPPGPAVGPAGERGEQAGSPGPGQGLPGPAGPGEAGKPGEG 600
QY 601 GVPDGLGAPGSPGAG 616
DB 601 GVPDGLGAPGSPGAG 616

RESULT 4
AAE02535
ID AAE02535 standard; protein; 1449 AA.
XX
AC AAE02535;

XX DT 10-AUG-2001 (first entry)
 XX DE Porcine alpha1(I) collagen.
 XX KW Porcine; alpha(I) collagen; gelatin; cytostatic; viral infection;
 KW pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;
 KW medical; arterial sealant; bone graft; dermal implant; cancer;
 KW rheumatoid arthritis; beverage; photographic application.
 XX OS Sus scrofa.
 XX FH Key Location/Qualifiers
 FT Misc-difference 929..830
 FT /note= "Encoded by ggcgaacctggtgatgctgctgtaaggcgtg
 FT ctggtcccccggccctgctgga"
 XX WO2001134647-A2.
 XX 17-MAY-2001.
 XX 10-NOV-2000; 2000WO-US030792.
 XX 12-NOV-1999; 99US-00439058.
 PR 10-NOV-2000; 2000US-00709700.
 XX (FIBR-) FIBROGEN INC.
 XX PI Bell MP, Neff TB, Polarek JW, Seeley TW;
 XX WPI; 2001-335911/35.
 DR N-PSDB; AAD06576.
 XX Novel isolated and purified bovine or porcine collagens and gelatins
 PT useful in medical, pharmaceutical, food and cosmetic industries, as
 PT vaccine, and for treating autoimmune disorders, infections and cancer.
 XX Example 3; Fig 8; 168pp; English.
 CC The present sequence is porcine alpha1(I) collagen. The present invention
 CC relates to recombinant synthesis of collagens and gelatins derived from
 CC animals. Collagen is useful in medical, pharmaceutical, food and cosmetic
 CC industries. Collagen is an important component of arterial sealants, bone
 CC grafts, drug delivery system, dermal implants, haemostats, and
 CC incontinence implants, and for treating autoimmune disorders such as
 CC rheumatoid arthritis. Collagen is useful in food products such as sausage
 CC casings, and in cosmetics or facial and skin products such as
 CC moisturisers. Recombinant gelatin is useful in vaccine formulations for
 CC treating viral infections, autoimmune diseases and cancer. Gelatin is
 CC useful in the manufacture or as a component of various pharmaceutical and
 CC medical devices and products, in food and beverage industries, in hair
 CC care and skin care products, as a glue or adhesive in various
 CC manufacturing processes, as a light-sensitive coating in various
 CC electronic devices, as photoresist base in photolithographic processes,
 CC in printing and photographic applications, in laboratory application, and
 CC as a component in various gels used for biochemical and electrophoretic
 CC analysis, including enzymographic gels
 XX Sequence 1449 AA;
 Query Match 59.6%; Score 2765.5; DB 4; Length 1449;
 Best Local Similarity 56.7%; Pred. No. 1.4e-162;
 Matches 559; Conservative 32; Mismatches 226; Indels 169; Gaps 21;
 QY 1 GPPGPGPTGLGPPGGRGFGS-----RGFFCA-----DGVAGPKGPAGRGSPGPA 48
 Db 193 GPPGAPGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 252
 QY 49 GPKSGPGGAGRP-----GEAGLPGAKLTGSPGSPGPDGKTGPPGAGQD 93
 Db 253 GARGLPATGLPMKMGHRSGLDAGKADGAPGPGPGPGPGPGPGPGPGPGPGPG 312
 QY 94 GRGPGPGPGARGQ---AGVMGPGPGKGAAGE---PKAGRGVPGPPGAVGPAKDGGA 147

Db 313 GRGPGPGAGARGNDGATGAAGPPGPGTGPAGPPGPGAVGAKGEAGPQARGSEGPQGV 372
 QY 148 GAQGPAGPAGPAGERGEQGP-----AGSEFGQGLPGPAGPPGEAKGPGQ 192
 Db 373 GEPGPPGAGAAAGPAGNPGADQPGKKGANGAPGAGPFGFARGPGSPGPGPGPK 432
 QY 193 GVPDGLGAP-----GPSGPA-----GEPGPTCLPGPPGERG 225
 Db 433 GNSGEPGAPGSKGDTGAKGEPGPTGVQGPFGPAGGEGKARGEPGAGLPGPGERG 492
 QY 226 GSRGFGADGVAGPKGPAGERSGPGPAGPKGSGFAGRPGGAGLPGCAKGLTSPSPGPD 285
 Db 493 GSRGFGADGVAGPKGPAGERSGPGPAGPKGSGFAGRPGGAGLPGCAKGLTSPSPGPD 552
 QY 286 GKTGPPGAGQDGRPGPPGPGCARGOAGVMGPGKGAAGBPKAGRGVGPFGPAGVGA 345
 Db 553 GKTGPPGAGQDGRPGPPGPGCARGOAGVMGPGKGAAGBPKAGRGVGPFGPAGVGA 612
 QY 346 GXDGSAGAQQPPGPGPAGERGEQGPAGSPGQGLPGPAGPPGEAKGPGGPGDLGAP 405
 Db 613 GKDGZAGAQQPPGPGPAGERGEQGPAGSPGQGLPGPAGPPGEAKGPGGPGDLGAP 672
 QY 406 GPSGPAGE-----PGP-----TGLPGPPGEGGSGSRGFP 435
 Db 673 GFSGARGRGFPGERGVQGPFGPAGPRGANGAPNDGAKGDAGAPGAPSGOAPLQMP 732
 QY 436 GADGVAGPKPAGERSGPGPAGPKGSGP-----AGRPGGAGLPGCAKGLTSPSPG 486
 Db 733 GERGAAGLPGPKGDRDAGPKGADGAPKDGVRGLTGPIGPPGAPAGPDKGETGSPGA 792
 QY 487 GPDGKTGPPGAGQDGRPGP---PGPPCARGOAGVMGF-----PGPKGAAGBPKAGRGV 539
 Db 793 GTTGARGAPDRGEPGPPGPGPAGFAGPPGADGQPGKAGGPTGPPGIVGVPAGPKGARG 852
 QY 540 PGPPCA-----VGPAGKDXGAGAOQPPGPGPAGPAGE---RGEQSGPAGSPGQGLPGP 587
 Db 853 AGPPGATGFPGAAGRVGPPGPGNAGPPGPGPAGKESKGRGETGPGAGRGPPG 912
 QY 588 AGPPGAGKPBQGVPGDLGAPGSPGAPGPGTGLPGPPGGERGPGSRGPPGADGVP 647
 Db 913 PGPAGEKSPGADGAPAGPAGTGPQGTAGQGVVGLPGQRGGERGPPGLPGSPGPPGQ 972
 QY 648 KGPAGERSGPGPKGSGFAGRPGGAGLPGCAKGLTSP---GSPGPD---GKTGPPG 701
 Db 973 SGPSGERGPPGMPG---PGLAGPPGSGRGAAGSGFORDGAPGKDRGSGGAPG 1029
 QY 702 AGQDGRPGPPGPGARGOAGVMGPPGPKGAAGBPKAGRGVGPFGPAGVAGPKDGE--- 758
 Db 1030 PGAPGAPGAPGVPVAGKSGDRGETGPGAPGAPGVPVAGRGAPGAPGQPRGDKGTGEGD 1089
 QY 759 -----AGAQGPPGP-----AGPAGERGEQGPAGSP---GFQGLPGPAGP 794
 Db 1090 RGKIHRSFGLQGPFGPPGPGSGGSGAPGPRGPGGAGAPGKDGKGLNGLPGPTG 1149
 QY 795 PGEAGKPGEQGVPGDLGAPGSPGAP 820
 Db 1150 PPRGRTGDAGVPVGPFGPPGPPGPPG 1175
 RESULT 5
 AAY84544
 ID AAY84544 standard; protein; 1057 AA.
 XX
 AC AAY84544;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE A human collagen 1 (alpha1) protein helical region.
 XX Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibrinogen; fibronectin; post translational hydroxylation.

XX OS Homo sapiens.
 XX FN EP992586-A2.
 XX XX
 XX PD 12-APR-2000.
 XX PF 07-OCT-1999; 99EP-00119184.
 XX PR 09-OCT-1998; 98US-00169768.
 XX PA (USSU) US SURGICAL CORP.
 XX PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX DR WPI; 2000-259138/23.
 XX DR N-PSDB; AAA12503.
 XX PT Production of extracellular matrix proteins containing 4-trans-
 XX PT hydroxyproline results in native self aggregating proteins, useful on
 XX PT medical implants.
 XX PS Example 10; Fig 39A-E; 260pp; English.
 XX CC The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents human collagen 1 (alpha1) helical region,
 XX which may be produced using the method of the invention
 XX SQ Sequence 1057 AA;

Query Match 59.6%; Score 2765; DB 3; Length 1057;
 Best Local Similarity 55.4%; Pred. No. 1.1e-162;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGPGTGLPGRGGRGSGRFPDAGVAGPKGAGRGSPGAPGKSGPEA--- 57
 DB 18 GPMGSGRGLPGRGGRGSGRFPDAGVAGPKGAGRGSPGAPGKSGPEA--- 57

QY 58 GRPGEAGLP--GAKLTSGSGSPG-----DGKTGPPGAGQDGRPGPPGPGAR 105
 DB 78 GRPGEAGLP--GAKLTSGSGSPG-----DGKTGPPGAGQDGRPGPPGPGAR 105

QY 106 GQAGVMGFPKGAAGEPKAGER-----GVPGFPGVAGPA-----GKQGEAGQ 150
 DB 138 GQAGVMGFPKGAAGEPKAGER-----GVPGFPGVAGPA-----GKQGEAGQ 150

QY 151 GPPGPGAPGAGEGEP-----AGSGPFGQLPGPA 180
 DB 198 GPPGPGAPGAGEGEP-----AGSGPFGQLPGPA 180

QY 181 GPPGPGAPGAGEGEP-----AGSGPFGQLPGPA 180
 DB 198 GPPGPGAPGAGEGEP-----AGSGPFGQLPGPA 180

QY 258 GPPGPGAPGAGEGEP-----AGSGPFGQLPGPA 180
 DB 258 GPPGPGAPGAGEGEP-----AGSGPFGQLPGPA 180

QY 214 GPPGPGAPGAGEGEP-----AGSGPFGQLPGPA 180
 DB 214 GPPGPGAPGAGEGEP-----AGSGPFGQLPGPA 180

QY 318 GLPFGPGRGGRGSGRFPDAGVAGPKGAGRGSPGAPGKSGPEAGLPGAK 377
 QY 274 GLTSGSPGSPGDPGKTGTPPGAGQDGRPGPPGAGQAGVMGFPKGAAGEPKAGER 333
 DB 378 GLTSGSPGSPGDPGKTGTPPGAGQDGRPGPPGAGQAGVMGFPKGAAGEPKAGER 437
 QY 334 GVPFGPAGVAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 393
 DB 438 GVPFGPAGVAGKDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 497
 QY 394 GEGVPGDLGAPGSPGAGE-----TGLPPOP 423
 DB 498 GEGVPGDLGAPGSPGAGE-----TGLPPOP 423

QY 424 GERGGSPGRGPPGAGVAGPKGAGRGSPGAPGKSGPE-----AGRPGEAGLP 474
 DB 558 GSQAGAPGAGQMPGEGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 617
 QY 475 GAK-----GLTSGSPGSPGDPGKTGTPPGAGQDGRPGPPGAGQDGRPGPPG 515
 DB 618 GAKSGSPGSPGDPGKTGTPPGAGQDGRPGPPGAGQDGRPGPPGAGQDGRPGPPG 677

QY 516 -----AGVMGFPKGAAGEPKAGRGVPGPPGA-----VGPAGKDGAGAGAGAG 561
 DB 678 GPPGPGAPGAPGPPGPIGNVAGPAGKAGAGAGAGAGAGAGAGAGAGAGAGAGAG 737

QY 562 GAGAP-----AGRGEGGSPGSPGDPGKTGTPPGAGQDGRPGPPGAGQDGRPGPPG 618
 DB 738 GAGKGGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 797

QY 619 GPTGLPGRGGRGSGRFPDAGVAGPKGAGRGSPGAPG---KGSPEAGRPGEA 675
 DB 798 GVVGLFGQGRGGRGSGRFPDAGVAGPKGAGRGSPGAPG---KGSPEAGRPGEA 857

QY 676 GLPAGKLTGSPGSPGDPGKTGTPPGAGQDGRPGPPGAGQDGRPGPPGAGQDGRPGPPG 735
 DB 858 GAGSGRGRGSGRFPDAGVAGPKGAGRGSPGAPG---KGSPEAGRPGEA 917

QY 736 GKAGRGVPGPAGVAGPAGKDG-----AGAGQSPG-----AGPA 771
 DB 918 GPKAGRGVPGPAGVAGPAGKDG-----AGAGQSPG-----AGPA 771

QY 772 GERGGEGGAGSP---GFGGLPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 820
 DB 978 GPRGPPGAG 1029

RESULT 6
 AAY84403
 ID AAY84403 standard; protein; 1058 AA.
 XX AAY84403;
 AC AAY84403;
 XX 12-JUL-2000 (first entry)
 DT 12-JUL-2000 (first entry)
 DE Amino acid sequence of human type 1 (alpha1) collagen polypeptide,
 KW Alpha1 collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
 KW collagen; mussel adhesive protein; bioadhesive.
 XX Homo sapiens.
 OS (USSU) US SURGICAL CORP.
 FN (PAOL) PAOLELLA D N.
 XX (GRUS) GRUSKIN E A.
 PD 16-MAR-2000.
 PF 07-SEP-1999; 99WO-US020462.
 XX 09-SEP-1998; 98US-0099652P.
 PR (USSU) US SURGICAL CORP.
 PA (PAOL) PAOLELLA D N.
 PA (GRUS) GRUSKIN E A.

PA (BUEC/) BUECHTER D D.
 PI Paoella DN, Gruskin EA, Buechter DD;
 XX WPI; 2000-271051/23.
 DR N-PSDB; AA299843.
 XX
 PT Incorporating non-natural amino acid into polypeptide, useful e.g. for
 PT production of bioadhesives, by epoxidation or substitution of
 PT dehydroproline residues.
 XX
 PS Disclosure; Fig 6; 66pp; English.
 XX
 CC The present sequence represents a human type 1 (alpha1) collagen protein.
 CC Peptides derived from the protein were used to demonstrate incorporation
 CC of 3,4-dehydro-L-proline into the peptide, using the method of the
 CC invention. The specification describes a method for the incorporation of
 CC non-natural amino acid into a polypeptide. The method comprises reacting
 CC at least one 3,4-dehydroproline residue in the polypeptide with an
 CC epoxidation reagent from a polypeptide containing at least one 3,4-
 CC epoxyproline residue. The method is used for studying the effects of non-
 CC natural amino acids on structure and function of polypeptides. The method
 CC is also useful for commercial production of collagen or mussel adhesive
 CC proteins (which are useful as bioadhesives), and for incorporating a wide
 CC variety of groups, including therapeutic ligands and biological probes,
 CC into polypeptides.
 XX
 SQ Sequence 1058 AA;

Query Match
 Best Local Similarity 59.6%; Score 2765; DB 3; Length 1058;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGPGPTGLPFPGERGSGSGFPGADGVAGPAGRGSPGAPGKSGPGEA--- 57
 Db 19 GFMGSGPRGLPFPGPAGPQGFQFPGEPGASGPMGPRGPPGPKNGDDGEAGKP 78
 QY 58 GPPGAGLP---GAKGLTSGSGSPG-----DKTGPAGAGQDGRPPGPPCAR 105
 Db 79 GRPGRGPPGQAGRLPGLTGLCMKGRHGFSLDGAKGADGAPGKSGPSPGNGAP 138
 QY 106 GQAGVMGPPGPKGAAGEPKAGER-----GVPPGPAVGPA-----GKDGGAQ 150
 Db 139 QMGPRGLPGRGRGPGAPGAPGAGNDGATGAAGPPTGAPPPGAVGAKGAGPQ 198
 QY 151 GPPGAGPAGRGSGGP-----AGSPFGQLFGPA 180
 Db 199 GPRGSEGPQVGRGPPGPPGAGAPGAGNPGADGQPKAGKANGAPGIAGAPFGARGPS 258
 QY 181 GPPGAGKPGQGVPGDLGAP-----GPGSPA-----GPPGPT 213
 Db 259 GPQGPFGPPGKNSGPGAPGSKDGTAKGEPGVGVGPPGPGFAGEGKRGARGEPGPT 318
 QY 214 GLPQPGERGPGSGRFPFGADGVAGPKPAGERSPGPAGPKSGPGEAGPGEAGLPKAK 273
 Db 319 GLPQPGERGPGSGRFPFGADGVAGPKPAGERSPGPAGPKSGPGEAGPGEAGLPKAK 378
 QY 274 GLTSGSGSPGDKGTGPPGAGQDGRPPGPGCARGQAGVMGFPKPGAAGPFGKAGER 333
 Db 379 GLTSGSGSPGDKGTGPPGAGQDGRPPGPGCARGQAGVMGFPKPGAAGPFGKAGER 438
 QY 334 GVPQPGAVGPAKDXGAGAGQGGPPGAPGAGEGQAGSPGQGLPGPAGPBGAGKP 393
 Db 439 GVPQPGAVGPAKDXGAGAGQGGPPGAPGAGEGQAGSPGQGLPGPAGPBGAGKP 498
 QY 394 GEQGVPGDLGAPGSGPAGE-----PGP-----TGJPGPP 423
 Db 499 GEQGVPGDLGAPGSGPAGEGPFGERGVQPGPAGPRGANGAPGNDGAKGDAGAPGAP 558
 QY 424 GERGGPSRGPFGADGVAGPKPAGERSPGPAGPKSGPGE-----AGSPGAGLP 474
 Db 559 GSCGAPLQGMWGERGAGLPFGPKGRDAGPKADGSKGQVRLTGPFGPPGAPAG 618

QY 475 GAK-----GLTSGSGSPGDKGTGPPGAGQDGRPPGPPGARGQ----- 515
 Db 619 GDKGESGSPGAGTGAAGAPGDRGPPGPPGAGFAGCQPKAGKGPAGDAGAKGDA 678
 QY 516 -----AGVMGFPKGAAGEPKAGERGVPPGPA-----VCPACKDGEAGAGQGP 561
 Db 679 GPPGAPGAGPPGPIGNVGAAPGAKGARGAGPPGATGFFCAAGRVGPPGSGNAGPPGP 738
 QY 562 GPAGP---AGERGEQAGSPGQGLPAGPAGPGEAGKPGQGVPGDLGAPGSPGAPGP 618
 Db 739 GPAGKEGKGPRGCTGPAAGPGEVPPGPPGAGEKSGFADGAPGAPCTPGQGIAGOR 798
 QY 619 GPTGLPQPGERGSGRFPFGADGVAGPKPAGERSGSPGAP---KSPGEAGRPGEA 675
 Db 799 GVVGLPQGRGEGFPGLPFGSGEPGKQSGASGERGPEGPMGPPGLAGPPGESREGAP 858
 QY 676 GLFGAKGLTSGSGSPGDPGKTGPPGAGQDGRPPGPPGARGQAGVMGFPKPKGAAGEP 735
 Db 859 GABGSGPRDGSFGAKGDRGCTGPAAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 918
 QY 736 GKAGERGVPPGPAAGVPAAGKAGE-----AGAQQPPGP-----AGPA 771
 Db 919 GPAGARGPAGPQGRDKGTGEGTGEQDGRGKGRHGFSLGQGGPPGPPGSGEAGSPGASGPA 978
 QY 772 GERGEQAGSP---GFOGLPAGPAGPGEAGKPGEGVPGDLGAPGSPGAP 820
 Db 979 GPRGPPSAGAPGKDLNGLPPIGPPGRGRTDAGVGVPPGPPGPPGPPG 1030

RESULT 7
 AA584540
 ID AA584540 standard; protein; 1107 AA.
 XX
 AC AA584540;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of a chimeric collagen 1 (alpha1)/decorin protein.
 DE Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibronectin; fibronectin; post translational hydroxylation;
 KW decorin; chimera.
 XX
 OS Homo sapiens.
 OS Unidentified.
 OS Chimeric.
 PH Key Location/Qualifiers
 FT Misc-difference 858
 FT /note= "Gly encoded by GCT"
 XX
 PN EP992586-A2.
 PD 12-APR-2000.
 XX
 PF 07-OCT-1999; 99EP-00119184.
 XX
 PR 09-OCT-1998; 98US-00169768.
 XX
 PA (USSU) US SURGICAL CORP.
 XX
 PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX
 DR WPI; 2000-259138/23.
 DR N-PSDB; AAA12500.
 XX
 PT Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX
 PS Claim 24; Fig 18; 260pp; English.
 XX

The specification describes a method for producing an extracellular matrix protein or its fragment. The extracellular matrix protein is capable of self aggregating in a cell which does not ordinarily hydroxylate prolines. The method comprises optimising a nucleic acid sequence for expression in the cell by substitution of codons preferred by that cell for naturally occurring codons not preferred by the cell; incorporating the nucleic acid sequence into the cell; and contacting the cell with a hypertonic growth medium containing at least one amino acid, selected from the group consisting of trans-4-hydroxyproline and 3-hydroxyproline to allow at least one of the amino acids to be assimilated into the cell and incorporated into the extracellular matrix protein. The method may be used to make host cells assimilate and incorporate trans-4-hydroxyproline into proteins. This is especially useful in the recombinant production of proteins such as collagen, fibronogen and fibronectin whose ability to self aggregate and produce functional proteins depends on the post translational hydroxylation of proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans-4-hydroxyproline. The present sequence represents a chimeric collagen 1 (alpha1)(I) decorin protein, which may be produced using the method of the invention

Sequence 1107 AA;
XX
SQ

Query Match 59.6%; Score 2765; DB 3; Length 1107;
Best Local Similarity 55.4%; Pred. No. 1.2e-162;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

[illegible]

CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a human protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1464 AA;

Query Match 59.6%; Score 2765; DB 7; Length 1464;
 Best Local Similarity 55.4%; Pred. No. 1.5e-162;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGFGTGLPGRGGRGSRFFPGADGVAGPKGAGRGSPGAGPKGSGEA--- 57
 DB 179 GPMGSGRGLPGRGAGPQGFQGPCEGPGAGPMGPRGPPGPKNGDDGAGKP 238
 QY 58 GEPGAGLP---GAKGLTSGSGSPG-----DGKTGPPGAGDGRPPGPPGAR 105
 DB 239 GRPGRGPPGPGARGLPCTAGLPQMKGHRGSLDCAKAGDAGPAGPKGSGSGENGAP 298
 QY 106 GQAGVMGPPGPKGAAGEPKGAGER-----GVPGPAGVGA-----GKDGAGAO 150
 DB 299 GQMGPRGLPGRGRGAGPAGAGNDGATCAAGPPGTGTPAGPPGAVGAKGAGPQ 358
 QY 151 GPPGAGPAGRGEGOP-----AGSPGFGPLGPA 180
 DB 359 GPRGSGPQVRGEPGPPGAGAGPAGNPGADGQPKAGKANGAPGIAGPFGPARGPS 418
 QY 181 GPPGAGKPGGQVPGDLGAP-----GPSGA-----GEPGPT 213
 DB 419 GPGGPGPPGPKGNSGEPGAGSGKDTGAKGEPGVGVGPPGAGBEGKRGARGEPGT 478
 QY 214 GLPGPPGRRGGRGPPGADGVAGPKGAGRGSPGAPGKSGPAGRPGGAGLPGAK 273
 DB 479 GLPGPPGRRGGRGPPGADGVAGPKGAGRGSPGAPGKSGPAGRPGGAGLPGAK 538
 QY 274 GLTSGSGSPGDKTGPGRAGQGRPPGPPGARGCAGVMGPPGKGAAGPFGKAGER 333
 DB 539 GLTSGSGSPGDKTGPGRAGQGRPPGPPGARGCAGVMGPPGKGAAGPFGKAGER 598
 QY 334 GVPFPAGVAGPKDAGAGAGPPGAPGAGERGEGQGPAGSPFGQLGPPAGPPGAGKP 393
 DB 599 GVPFPAGVAGPKDAGAGAGPPGAPGAGERGEGQGPAGSPFGQLGPPAGPPGAGKP 658
 QY 394 GEGVFGDLGAPSGPAGE-----PGP-----TGLPGPP 423
 DB 659 GEGVFGDLGAPSGPAGE-----PGP-----TGLPGPP 718
 QY 424 GERGGPSRSPGADGVAGPKGAGRGSPGAPGKSGPGE-----AGRPGAGLP 474
 DB 719 GSQAGPLQMPGRRGAAGLPFGPKDGDGAGPKGADSGPKDVGRLTGPPIGPPGAGLP 778
 QY 475 GAK-----GLTSGSGSPGDKTGPGRAGQGRPPGAPGPPGPPGARGO----- 515

DB 779 GDKGESGSPGAGTGAAGAPGDRGPPGPPGAGFAGFCADGQCGKAGKPGDAGAKGDA 838
 QY 516 -----ACVMGFPKPKGAAGSPGKAGRGVGPBPPCA-----VGPAGKDGAGAGQPP 561
 DB 839 GPPGAGPAGPPGPIGNVGNAPGAKGARGAGPPGATGFFCAAGRVGPPGSGNAGPPGP 898
 QY 562 GPAGP---AGERGEGQGPAGSPGQGLPGPAGPPGKAGKPGEGQVPGDLGAPGSPGAPG 618
 DB 899 GPAGKGGKGRGETGAPRPGVGPVGPAGKSGSPCADGPAGAPGTGPGQGIAGQR 958
 QY 619 GPTGLPGRGGRGSPGSRGFPAGDGVAGPKGAGRGSGPAGP---KSGPAGRPGEA 675
 DB 959 GVVGLPGQGRGFPGLPGSPGEPGKQSGSGASGERGPPGPMGPPGLAGPPGSGREGAP 1018
 QY 676 GLPGAKGLTSGSPGSPGDKTGTTPGAGQDGRPPGPPGARGQAGVMGFPKGAAGBP 735
 DB 1019 GAEZGSPDRGSPGAKGDRGETGAPGPPGAPGAPGPPVGPAGKSDRGETGAPGAPPV 1078
 QY 736 GKAGRGVGPAGPAGVAGPAGKAGE-----AGAGPPGP-----AGPA 771
 DB 1079 GPAGARGPAGPQGRDKGETGEGQDRGKTKHGRGSLGQGPFGPPGSGEGQGPASGPA 1138
 QY 772 GERGEGQGPAGSP---GFGQLPGPAGPPGAGKPGGQGVGPDLCAGPSPGAPG 820
 DB 1139 GPRGPPGSAGAPGKDLGGLGPIGPPGPRGRTGAGPVGPPGPPGPPGPPG 1190

RESULT 11
 ADD45055
 ID ADD45055 standard; protein; 1464 AA.
 XX AC ADD45055;
 XX DT 29-JAN-2004 (first entry)
 XX DE Human Protein P02452, SEQ ID NO 10487.
 XX KW Human; pain; neuronal tissue; gene therapy;
 XX KW spinal segmental nerve injury; chronic constriction injury; CCI;
 XX KW spared nerve injury; SNI; Chung.
 XX OS Homo sapiens.
 XX PN WO2003016475-A2.
 XX PD 27-FEB-2003.
 XX PF 14-AUG-2002; 2002WO-US025765.
 XX PR 14-AUG-2001; 2001US-0312147P.
 XX PR 01-NOV-2001; 2001US-0346382P.
 XX PR 26-NOV-2001; 2001US-0333347P.
 XX PA (GEHO) GEN HOSPITAL CORP.
 XX PA (FARB) BAYER AG.
 XX PI Woolf C, D'urso D, Befort K, Costigan M;
 XX DR WPI: 2003-268312/26.
 XX DR GENBANK; P02452.
 PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 PS Claim 1; Page; 1017pp; English.

CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent

CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1464 AA;

Query Match 59.6%; Score 2765; DB 7; Length 1464;
Best Local Similarity 55.4%; Pred. No. 1.5e-162;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGEPGTLGPPGGRGGSGFPGADGAGVAGPKGAGRGSPGAGKSPGEA--- 57
DB 179 GPMGSGRGLPGLPPGAPGPGFPGGPEGPGASGPMGRGPPGKNGDDGAGKP 238
QY 58 GRPGEAGLP---GAKGLTSGSPGSP-----DGKTGPPGAGDGRPPGPPGAP 105
DB 239 GRPGERGPPGQAGRLPGTAGLPMKGRHGFSLDGAKDAGAGPKGPGSPGNGAP 298
QY 106 QGAVMGPPGPKGAAGRGKAGER-----GVPPPGAVGPA-----GKDGGAQA 150
DB 239 QCMGPRGLPGRGPGAPGPGAGRGNDGATGAAGPPOPTGAGPPGPFVAGKAGAPQ 358
QY 151 GPPGAPGAGRGEGP-----AGSPGFQGLGPA 180
DB 359 GPRGSEGGVGRGFPGPPAGAGPAGNPGADQCPGAKGANGAGPIAGAPGFGARGPS 418
QY 181 GPPGAGKPGGQGVGDLGAP-----GPSGPA-----GEGPPT 213
DB 419 GPQGGPGPGPKGNSGEPGAPGSKGDTGAKGEPFVGQPPGAGEGKRGARGEFPT 478
QY 214 GLPGRGGRGGSGFPGADGAGVAGPKGAGRGSPGAGKSPGAGRPGEAGLPCA 273
DB 479 GLPGRGGRGGSGRFPAGDAGVAGPKGAGRGSPGAGKSPGAGRPGEAGLPCA 538
QY 274 GLTSGSPGPPDGKTGPPGAGDGRPPGPPGARGQAGVGMGPPGKGAAGEGKAGER 333
DB 539 GLTSGSPGPPDGKTGPPGAGDGRPPGPPGARGQAGVGMGPPGKGAAGEGKAGER 598
QY 334 GVPGPAGVAGPKDGEAGAQPPGAPGAGRGEGQAGSPGQGLPGAPGAGEAGKP 393
DB 599 GVPGPAGVAGPKDGEAGAQPPGAPGAGRGEGQAGSPGQGLPGAPGAGEAGKP 658
QY 394 GEGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
DB 659 GEGVPGDLGAPGSPGAGERGFPGERGVQPPGPPGANGAPGNDGAGDAGAP 718
QY 424 GERGGSGRGGFPGADGAGVAGPKGAGRGSPGAGKSPGGE-----AGRGEGALP 474
DB 719 GSGAGPLQCMFGERGAAGLPKGDGAGPKGADGSGFGKDGVRGLTGTPTGPPGAP 778
QY 475 GAK-----GLTSGSPGPPDGKTGPPGAGDGRPPGPPGAPGQ----- 515
DB 779 GDRGSGSGPPGAPGTGARGAPDRGEPGPPGAPGAPGADGQCGAKGEGDAGAKDA 838
QY 516 -----AGVMGFPKPGKGAAGEGPKAGERGVGPPGPA-----VGPAGKGEAGQAPP 561

DB 839 GPPGAPGAGPPGPTGNVAGFAGKAGSAGPPGATGPPGAGRVGPPGSGNAGPPGPP 898
QY 562 GPAGP---AGRGSGGAGSGFQGLPGAPGPPGAGKPGQGVGDLGAPGSPGAPG 618
DB 899 GPAGKGGKGRGETGAPRGEGVGPFPAGPKGSGADGAPAGTGTGPGQIAGQR 958
QY 619 GPTGLGPPGGRGGSGRFPFADGAGVAGKPGAGRGSPGAPG---KGSFGEAGRPGEA 675
DB 959 GVVGLPGQGRGGRGFPGLPGSGEPKQGFSGASGERGPPGPMGPPGLAGPFESGREGAP 1018
QY 676 GLPGAAGLTGSPGSGPPGKTPGAGDGRPPGPPGARGQAGVGMGPPGKGAAGEP 735
DB 1019 GAESGFGDGGSGAKGDRGTGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPV 1078
QY 736 GXAGRGVGPVGPAGVAGPKDGE-----AGAQQPPGP-----AGPA 771
DB 1079 GPAGARGGAPGQPRGDKGTGEGQDGRGKGRHGFSLGQPPGPPGSPGEGQPSGASGA 1138
QY 772 GRRGQGGAGSP---GFGGLGAPGAPGAGKPGQGVGDLGAPGSPGAPG 820
DB 1139 GPRGFPGSAGAFKDGGLNGLPFIIGPGRGTGDPGVPVGPFPGPPGPPG 1190

RESULT 12
ADD45051
ID ADD45051 standard; protein; 1464 AA.
XX
AC ADD45051;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human Protein P02452, SEQ ID NO 10483.
XX
KW Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WFI; 2003-268312/26.
XX
DR GENBANK; P02452.
XX
PT New composition comprising two or more isolated polypeptides, useful for
FT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially

expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (thung), chronic constriction injury (CCI), and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at www.wipo.int/pub/published_pct_sequences.

Sequence 1464 AA;

[illegible]

CC fraction, obtained from deer antler cartilage cells, for inhibiting cell
CC growth and/or division by inserting into an animal cell, a compound which
CC inhibits the translation of the polynucleotide encoding the DACC. The
CC method and the polypeptides are useful for stimulating mesenchymal cell
CC growth and/or division or for stimulating chondrogenesis, cartilage, disc
CC or connective tissue growth, repair, regeneration and/or restoration in
CC an animal. The polynucleotides, polypeptides, agonists and antagonists
CC may be used in treatment modalities, specifically in gene therapy. The
CC polypeptides can be used as bait proteins in a two- or three-hybrid assay
CC to identify other proteins, which bind to or interact with the
CC polypeptide and are involved in modulating cell growth and/or division.
CC The sequences presented in ABG93923-ABG93948 are the proteins encoded by
CC the DACC cDNA clones
XX
XX
SQ Sequence 1461 AA;

Query Match 59.5%; Score 2761; DB 5; Length 1461;
Best Local Similarity 55.3%; Pred. No. 2.6e-162;
Matches 560; Conservative 31; Mismatches 229; Indels 192; Gaps 19;
QY 1 GPPGEPGTLGPPGGRGPGGRRGFGAGDGVAGPKPAGERSPPGAGKSPGEA--- 57
DB 176 GPMGPGRGLPGPPGAGPGQFGPGGEPGCEGFCASGPMGPRGPPGKXNGDDGAGKP 235
QY 58 GRPCEAGLP---GAKGLTSPGSPG---DCKTGPAGQDGRPPGPPGPGAR 105
DB 236 GRPGERPPGQAGRLPPTAGLDMKCHRGFSLDCAKDGADGAPGKGFSPGENGAP 295
QY 106 GQAGVMGPPGPKGAAGFPKAGER-----GVPPGPAVGA---GKDGAGAG 150
DB 296 GQMGPRGLPGRGRGPAAGPAGARGNDGATGAAGPPGTGAPGPPGPAVGAAGAGPQ 355
QY 151 GPPGAPGAGRGQGP-----AGSPGFGQLGPA 180
DB 356 GPRGSEGGVGRGPPGPPGAGAGPAGNPGADGQPKAGANGAPGAGPFGGARGPS 415
QY 181 GPPGAGKPGQGVFGDLGAP-----GPGSPA-----GEPGPT 213
DB 416 GPQFGPPGPKNSGFBGAGSKDGTGAKGPGVGVQGGPPGAGEGKRGARGEPGT 475
QY 214 GLPGRGGRGGRGPPGAGDGVAGPKPAGERSPPGAPKSPGAGRGAGLPGAK 273
DB 476 GLPGRGGRGGRGPPGAGDGVAGPKPAGERSPPGAPKSPGAGRGAGLPGAK 535
QY 274 GLTSPGSPGPDGKTGTPGAGQDGRPPGPPGARGQAGVWGPFGKGAAGEPKAGER 333
DB 536 GLTSPGSPGPDGKTGTPGAGQDGRPPGPPGARGQAGVWGPFGKGAAGEPKAGER 595
QY 334 GVPGPAGVAGPKDGEAGAGQPPGAPGAGERGQGPAGSPGQGLPGPAGPGEAGKP 393
DB 596 GVPGPAGVAGPKDGEAGAGQPPGAPGAGERGQGPAGSPGQGLPGPAGPGEAGKP 655
QY 394 GQGVPGDLGAPGSPGAGE-----RGP-----TGLPGPP 423
DB 656 GQGVPGDLGAPGSPGAGERGPFGERGVQGGPPGPRGANGAPGNDGAKGDAGAP 715
QY 424 GERGGGSRGFFGAGDVAGPKPAGERSPPGAPKSPGFE-----AGRPGAGLP 474
DB 716 GSQGAPGLQMGFGERGAGLPGPKDGRDAGPKGADGSGKDXGVRGLTPIGPPGAPAG 775
QY 475 GAK-----GLTSGSPGPDGKTGTPGAGQDGRPPGPPGARGQ----- 515
DB 776 GDKGSGSPGAPGTGARGAPDGRGPPGPPGAPGAPGADGQGKAGEPGDAGAKGDA 835
QY 516 -----AGVMGFPKGAAGEPKKAGERGVGPPGPA-----VGPAGKDGAGAGQPP 561
DB 836 GPPGAPGAPGPPGTGTVGAPAKKARSAGPFGATGFGAAGRVGPPGSPGNAGPPGPP 895
QY 562 GPAGP-----AGERGEGPAGSPQGLPGPAGPFGAGKPGBOGVPGDILGAPGSPGAGEP 618
DB 896 GPAGKEGKGRGTGAPRCEVGPVPPGPPGAGEKSPGADGAPAGPCTGPGQIAGQR 955
QY 619 GTGLPGPPGREGGSGRFFGADGVAGPKPAGERSGSPGAPG---KGSPEGACRPGEA 675

DB 956 GVVLPGQGRGRRGFFGLPGSPGEGKQKQPSGASGERGPPGPMWPGLAGPFGESGREGAP 1015
QY 676 GLPGAKGLTSPGSPGPDGKTPPGPAGQDGRPPGPPGARGQAGVWGFPGPKAGAGEP 735
DB 1016 GAEGSPGRDGSFGAKGDRGETGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 1075
QY 736 GKAGERGVGPPGAVGAPGAKDGE-----AGAQQPPGP-----AGPA 771
DB 1076 GPGVARGPAGPQGGPKGKTGEGQDRIKGRGFGSLQGGPPGPPGSPGEOGSPGASGPA 1135
QY 772 GERGQGPAGSP---GFGQLEPGAPGPPGEGAKGQGVGVDLGAFFGSPGAPG 820
DB 1136 GPRGPPGSAGAPGKDLGLPGPIGPPGPRGRTGADGVPVGPFPFPPGPPG 1187

RESULT 14
AAU14136
ID AAU14136 standard; protein; 1464 AA.
XX
AC AAU14136;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human novel protein #7.
XX
KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytotstatic; neuroprotective; vulnerary; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX
OS Homo sapiens.
XX
FN WO200155437-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US002623.
XX
PR 25-JAN-2000; 2000US-00491404.
XX
PA (HYSE-) HYSE INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
PI WPI; 2001-451939/48.
XX
PT N-PSDB; AAS22441.
XX

Isolated polypeptides useful for treating anti-inflammatory diseases,
nervous system disorders, and for regenerating bone and cartilage.
Example 4; Page 525-527; 894pp; English.

The invention relates to polynucleotides encoding novel human proteins or
their active domains. The polypeptides, polynucleotides and antibodies
raised against the polypeptides are used in a method of treatment of a
mammal and prevention of disorders caused by the aberrant protein
expression or activity. The polypeptides can be used as molecular weight
markers, food supplements, and in antibody production. The polypeptides
are used to identify compounds which bind to the polypeptides.
Polynucleotides of the invention are used as probes and primers, for
sequencing, for chromosome or gene mapping, in the production of
recombinant proteins, and in generating anti-sense DNA or RNA and in gene
therapy. Polypeptides of the invention can be used to target drugs to a
tumour, in assays to determine biological activity, to raise
antibodies/ elicit an immune response, to determine quantitative protein
levels, as tissue markers, and to isolate receptors or ligands.
Polypeptides of the invention may also be useful in treating platelet
disorders, stem cell disorders, regenerating bone, cartilage, tendon,
ligament and/or nerve tissue, wound healing, treating burns, promoting

CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
 CC diseases, nervous system disorders, and infection. The present sequence
 CC represents a protein of the invention
 XX
 XX Sequence 1464 AA;

Query Match 59.5%; Score 2761; DB 4; Length 1464;
 Best Local Similarity 55.3%; Pred. No. 2.6e-162; Indels 192; Gaps 19;
 Matches 560; Conservative 31; Mismatches 229;
 QY 1 GPPGPGTGLPFPGERGSGRFPFGADGVAGPKGAPGERGSPGAPGKSGPCEA--- 57
 DB 179 GFMGSPGRGLPFPAGPQGFQGPGEPCGASGPMGPRGPPGPKNGDDGEAGKP 238
 QY 58 GPGGAGLP---GAKLITGSGSPCP-----DGKTGPPGAGQDRPPGPPGAR 105
 DB 239 GRPGERGPPGQAGRLPGLTGLPKMKHGRGSLGDKAGDAGPAGPKGEPGSENGAP 298
 QY 106 GGAGVMGPPGKGAAGBEGKAGER-----GVPPGPGAVGPA-----GKDGAGAG 150
 DB 299 GQMGRGLPGRGRFGAPGAPAGRGNDATCAAGPPTGTPGPPGPGVAGKAGEAGCP 358
 QY 151 GPPGAPGAPGERGEGP-----AGSPFGQGLPGPA 180
 DB 359 GPRSEGGQVGRGPPGPPGAGAGPAGNPGADQPGKAGKANGAGPAGPFGPARGPS 418
 QY 181 GPPGAGKPGQGVGDLGAP-----GPGSGPA-----GERGPT 213
 DB 419 GPQPGGPPGKNGSEFAGKSGDGTAKGEPDPVGVQPPGABEGKRGARBPPT 478
 QY 214 GLPFPGERGPPGPPGADGVAGPKGAPGERGSPGAPGKSPGAGPFGAGLPGAK 273
 DB 479 GLPFPGERGPPGPPGADGVAGPKGAPGERGSPGAPGKSPGAGPFGAGLPGAK 538
 QY 274 GLTSGSPGPDGKTGPPGAGQGRPPGPPGARGQAGVMGPPGPKGAAGPFGKAGER 333
 DB 539 GLTSGSPGPDGKTGPPGAGQGRPPGPPGARGQAGVMGPPGPKGAAGPFGKAGER 598
 QY 334 GVPFPAGVAGKDXGAGAGPFPAGPAGERGQGPAGSPFGQLGPPAGPFGAGKP 393
 DB 599 GVPFPAGVAGKDXGAGAGPFPAGPAGERGQGPAGSPFGQLGPPAGPFGAGKP 658
 QY 394 GEQGVGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
 DB 659 GEQGVGDLGAPGSPGARGRFPGERGVQGPFPAGPRGANGAPGNDGAKGADAGAPGAP 718
 QY 424 GERGGPSRFPFGADGVAGPKGAPGERGSPGAPGKSPGAGE-----AGSPGAGLP 474
 DB 719 GSQAGPLQGWGGERGAGLPKPKGDRDAGPKGADGSPKGVRLTPIGPPGAPGAP 778
 QY 475 GAK-----GLTSGSPGPDGKTGPPGAGQGRPPGPPGARGQ----- 515
 DB 779 GDKGESGFPAGPTGARGAPGDRGEPGPPGAPGADGQCPKAGKGPAGKAGDA 838
 QY 516 ----AGVMGFPFGKGAAGPFGKAGERGVPPGPA-----VGPAGKDXGAGQGP 561
 DB 839 GPPGAPGAPGPPGIVNGVAGKAGAGSAGPPGATGFGCAAGRVGPPGSGNAGPPGPP 898
 QY 562 GPAGP---AGERGEQGPAGSPFGQLGPPGAPGPEAGKPGEQGVGDLGAPGSPGAPGP 618
 DB 899 GPAGKEGKGRGTGTPAGPAGVGVPPGPPGAPGKSGPGADGAPAGPTGPPQGIAGOR 958
 QY 619 GPTGLPFPGERGPPGPPGADGVAGPKGAPGERGSPGAPG---KSPGAGPAGP 675
 DB 959 GVVGLPQGRGERGPPGLPFGSGEPGKSGGSGRGPFGPPGPPGLAGPPGESREGAP 1018
 QY 676 GLPAGKGLTSGSPGPDGKTGPPGAGQDRGPPGPPGARGQAGVMGFPFGKGAAGBP 735

DB 1019 GAEGSPRDSGPGKAGDRGTGTGAPGPPGAPGAPGPPVPGKSGDRGTGTGAPGAPV 1078
 QY 736 GKAGERGVGPPGAVCPACKDGE-----AGAGGPPGP-----AGPA 771
 DB 1079 GPVAGRPPAGPPGPRDKGTGETEQDGRGKHKHGRGSLGQGPFPFGSGEQGSGASGPA 1138
 QY 772 GERGEQGPAGSP---GFOGLPFPAGPPGAGKPGQGVGPDLCARPSPGAPG 820
 DB 1139 GPRGPPGSAGAPGKDLGNGLPFGIPGPPGPRGTGAGPVGPPGPPGPPG 1190

RESULT 15

AAV84541

ID AAV84541 standard; protein; 1057 AA.

XX AAY84541;

DT 25-JUL-2000 (first entry)

XX Amino acid sequence of a human collagen 1 (alpha1) protein.

XX Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibrinogen; fibronectin; post translational hydroxylation.

XX Homo sapiens.

XX EP992586-A2.

XX 12-APR-2000.

XX 07-OCT-1999; 99EP-00119184.

XX 09-OCT-1998; 98US-00169768.

XX (USU) US SURGICAL CORP.

XX Gruskin EA, Buechter DD, Zhang G, Connolly K;

XX WPI; 2000-259138/23.

DR N-PSDB; AAA12502.

PT Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.

PS Disclosure; Fig 27A-E; 260pp; English.

CC The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a human collagen 1 (alpha1) protein, which
 CC may be produced using the method of the invention

SQ Sequence 1057 AA;

Query Match

Best Local Similarity 59.5%; Score 2759; DB 3; Length 1057;

Pred. No. 2.7e-162;

Search completed: October 18, 2004, 13:37:32
Job time : 135.602 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:31:20 ; Search time 32.5431 Seconds
(without alignments)
1673.075 Million cell updates/sec

Title: US-10-658-989A-3
Perfect score: 4640
Sequence: 1 GPGGPGPTGLPDPGRRGG.....GEGVFGDLGAPGSPGAG 821

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2765	59.6	1464	4	US-09-331-347C-21
2	2752	59.3	1461	4	US-09-585-887-9
3	2752	59.3	1461	4	US-09-289-578-9
4	2746	59.2	1057	3	US-08-931-820-1
5	2704	58.3	1341	3	US-08-963-825-18
6	2704	58.3	1341	3	US-09-500-811-18
7	2704	58.3	1341	3	US-09-570-573-18
8	2704	58.3	1341	3	US-09-548-608-18
9	2511	54.1	822	3	US-09-219-849-49
10	2488	53.6	1017	4	US-08-468-996-10
11	2478	53.4	1064	1	US-08-642-255-62
12	2476	53.4	1060	3	US-08-931-820-3
13	2476	53.4	1418	3	US-08-963-825-20
14	2476	53.4	1418	3	US-09-010-999-1
15	2476	53.4	1418	3	US-09-500-811-20
16	2476	53.4	1418	3	US-09-570-573-20
17	2476	53.4	1418	3	US-09-548-608-20
18	2464	53.1	1442	5	PCT-US95-02251-12
19	2464	53.1	1442	5	US-08-316-650-12
20	2370	51.1	1057	3	US-08-931-820-4
21	2350.5	50.7	1078	3	US-08-963-825-21
22	2350.5	50.7	1078	3	US-09-500-811-21
23	2350.5	50.7	1078	3	US-09-570-573-21
24	2350.5	50.7	1078	3	US-09-548-608-21
25	2267.5	48.9	1366	3	US-08-963-825-19
26	2267.5	48.9	1366	3	US-09-500-811-19
27	2267.5	48.9	1366	3	US-09-570-573-19

28 2267.5 48.9 1366 3 US-09-548-608-19
29 2265.5 48.8 1366 4 US-09-585-887-10
30 2265.5 48.8 1366 4 US-09-289-578-10
31 2255.5 48.6 1024 3 US-08-931-820-2
32 2254.5 48.6 1065 1 US-08-642-255-72
33 2169 48.7 1806 4 US-09-919-497-56
34 2156.5 46.5 960 3 US-09-219-849-5
35 2149.5 46.3 777 1 US-08-642-255-53
36 2144.5 46.2 720 3 US-09-219-849-4
37 2123.5 45.8 897 1 US-08-397-633A-50
38 2096 45.2 595 3 US-09-219-849-48
39 2096 45.2 595 3 US-09-219-849-50
40 2057 44.3 1745 4 US-09-795-061-4
41 2055 44.3 837 1 US-08-175-155-68
42 2055 44.3 837 1 US-08-477-509B-103
43 2055 44.3 837 1 US-08-642-255-101
44 2055 44.3 837 2 US-08-707-237A-75
45 2055 44.3 837 3 US-08-482-085B-103

RESULT 1
US-09-331-347C-21
; Sequence 21, Application US/09331347C
; Patent No. 6617431
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants,
; TITLE OF INVENTION: Obtaining Such and Their Uses
; FILE REFERENCE: 1149-3
; CURRENT APPLICATION NUMBER: US/09/331,347C
; CURRENT FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 21
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-331-347C-21

ALIGNMENTS

Query Match 59.6%; Score 2765; DB 4; Length 1464;
Best Local Similarity 55.4%; Pred. No. 2.9e-171;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPGGPGPTGLPDPGRRGGPPGADGVAGPKGAGRGSPGPKGSPGGEA--- 57
DB 179 GPMGSGPRGLPDPGAPGPGFQGPGEPPGSGPMGPRGPPGKNGDDGEAGKP 238
QY 58 GRPGEAGLP---GAKGLTSGSPGP-----DGKTGPPGAGQDGRPPGPPGAR 105
DB 239 GRPGERGPPGQARGLPOTAGLPGMKHGRGSLDGAKGADGAPGPKGPGSPGNGAP 298
QY 106 GQGVNGVFPKGKAGEPKGAGER-----GVPPGPAVGPA-----GKDEAGNQ 150
DB 299 GQMGPRGLPGERGPRGAPGAGNDGATGAAGPPGTPGPPGPAVGAKEGAGGQ 358
QY 151 GPGGPAAGPAGERGEOGP-----AGSPGFQGLPGPA 180
DB 359 GRFGSEPGVKGEPGPPGPAAGAGNPGADGPGAKGANGAGIAGAPFPGARGPS 418
QY 181 GPPGEAGKPGEGVPGDLGAP-----GPSGA-----GEPGPT 213
DB 419 GPGGPGPPGKNGSGEPGAGSKGDTAKGEPGVGVQGPAGGEGKRGARGEPGPT 478
QY 214 GLPGRGERGSGSGFPAGDGVAGPKGAGRGSPGPKGSPGCEAGRPGCEAGLPKAK 273
DB 479 GLPGRGERGSGSGFPAGDGVAGPKGAGRGSPGPKGSPGCEAGRPGCEAGLPKAK 538
QY 274 GLTSGPSGPPGPKTTPPGAPGQDGRPPGPPGPPGAGQAGVNGFFPGPKGAAGPGKAGR 333

Db 539 GLTSGSPGPDGKGTGPPGPPAGQDGRPPGPPGARGQAGVMGPPGKGAAGFPGKAGER 598
Qy 334 GVPPPGAVGAGKDGAG 393
Db 599 GVPPPGAVGAGKDGAG 658
Qy 394 GEQVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
Db 659 GEQVPGDLGAPGSPGAGEGFFGERGVQVPPGPPGARGANGAGAGAGAGAGAGAGAG 718
Qy 424 GERGGGSRGPPGADGAG 474
Db 719 GSQAGPLQGNFGRGAG 778
Qy 475 GAK-----GLTSGSPGPDGKGTGPPGPPAGQDGRPPGPPGARGO----- 515
Db 779 GDKGESGSPGAGPTGARGAGPDGRGPPGPPGAGAGAGAGAGAGAGAGAGAGAGAGAG 838
Qy 516 -----AGVMGPPGKGAAGFPGKAGERGVPGCCA-----VGPAGKDGAGAGAGPP 561
Db 839 GPPGAG 898
Qy 562 GPAGP-----AGRGSGGAG 618
Db 899 GPAGKGGKGRGTGAG 958
Qy 619 GPTGLPGRGPPGSRGPPGADGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 675
Db 959 GVVGLPQGRGPPGGLPGSGGPPGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1018
Qy 676 GLPAGKGLTSGSPGPDGKGTGPPGPPAGQDGRPPGPPGARGQAGVMGPPGKGAAGP 735
Db 1019 GAEGSPGDSGKAGDGRGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1078
Qy 736 GKAGERGVPAGGAG 771
Db 1079 GPAGARGAG 1138
Qy 772 GEREGQAGAGP-----GFGGLPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 820
Db 1139 GPRPPGAG 1190

RESULT 2

US-09-585-887-9
; Sequence 9, Application US/09585887
; Patent No. 6413742
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/585,887
; CURRENT FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/289,578
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-585-887-9

Query Match 59.3%; Score 2752; DB 4; Length 1461;
Best Local Similarity 55.2%; Pred. No. 2e-170;

Matches 559; Conservative 31; Mismatches 230; Indels 192; Gaps 19;
Qy 1 GPPGPGTGLPGLPGRGSGRGGFPGADGAGVAGPKPAGGERGSPGAGPKGSPCEA--- 57
Db 176 GPMGSGRGLPGLPGRGSGRGGFPGADGAGVAGPKPAGGERGSPGAGPKGSPCEA--- 235
Qy 58 GRPGAGPL---GAKGLTSGSPGSPG-----DKTGPAGAGQDGRPPGPPGARG 105
Db 236 GRPGAGPL---GAKGLTSGSPGSPG-----DKTGPAGAGQDGRPPGPPGARG 295
Qy 106 GQAGVMGPPGKGAAGFPGKAGER-----GVPGPPGAGVGA-----GKDGAGAG 150
Db 296 GQAGVMGPPGKGAAGFPGKAGER-----GVPGPPGAGVGA-----GKDGAGAG 355
Qy 151 GPPGAG 180
Db 356 GPRGSGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 415
Qy 181 GPPGAG 213
Db 416 GPPGAG 475
Qy 214 GLPGLPGRGPPGSRGPPGADGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 273
Db 476 GLPGLPGRGPPGSRGPPGADGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 535
Qy 274 GLTSGSPGPDGKGTGPPGPPAGQDGRPPGPPGARGQAGVMGPPGKGAAGFPGKAGER 333
Db 536 GLTSGSPGPDGKGTGPPGPPAGQDGRPPGPPGARGQAGVMGPPGKGAAGFPGKAGER 595
Qy 334 GVPGPPGAVGAGKDGAG 393
Db 596 GVPGPPGAVGAGKDGAG 655
Qy 394 GEQVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
Db 656 GEQVPGDLGAPGSPGAGEGFFGERGVQVPPGPPGARGANGAGAGAGAGAGAGAGAG 715
Qy 424 GERGGGSRGPPGADGAG 474
Db 716 GSQAGPLQGNFGRGAG 775
Qy 475 GAK-----GLTSGSPGPDGKGTGPPGPPAGQDGRPPGPPGARGO----- 515
Db 776 GDKGESGSPGAGPTGARGAGPDGRGPPGPPGAGAGAGAGAGAGAGAGAGAGAGAGAG 835
Qy 516 -----AGVMGPPGKGAAGFPGKAGERGVPGCCA-----VGPAGKDGAGAGAGPP 561
Db 836 GPPGAG 895
Qy 562 GPAGP-----AGRGSGGAG 618
Db 896 GPAGKGGKGRGTGAG 955
Qy 619 GPTGLPGRGPPGSRGPPGADGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 675
Db 956 GVVGLPQGRGPPGGLPGSGGPPGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1015
Qy 676 GLPAGKGLTSGSPGPDGKGTGPPGPPAGQDGRPPGPPGARGQAGVMGPPGKGAAGP 735
Db 1016 GAEGSPGDSGKAGDGRGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1075
Qy 736 GKAGERGVPAGGAG 771
Db 1076 GPRPPGAG 1135
Qy 772 GEREGQAGAGP-----GFGGLPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 820
Db 1136 GPRPPGAG 1187

RESULT 3
US-09-289-578-9

```
; Sequence 9, Application US/09289578
; Patent No. 6428978
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/289,578
; CURRENT FILING DATE: 1999-04-10
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-289-578-9

Query Match 59.3%; Score 2752; DB 4; Length 1461;
Best Local Similarity 55.2%; Pred. No. 2e-170;
Matches 559; Conservative 31; Mismatches 230; Indels 192; Gaps 19;

QY 1 GPPGEPPTGLPFPGRGSGRFGADGAGVAGKGPAGERSGPGAGKGPGEA--- 57
DB 176 GPMFGSPRGLPFPFAGPQGFQPGEPCEGAFMGPRGPPGPKNGDDGEAGK 235
QY 58 GRPEAGL---GAKGLTSGSGSP-----DGKTGPPGAGODRPPGPPGAR 105
DB 236 GRPGERPPGQAGRLPFTAGLPMKMRHGFSLDGAKDAGPAGKSGFSGENGAP 295
QY 106 QGAGVMGFPKGAAGPGKAGER-----GVPPGPAVGA-----GKDGEAG 150
DB 296 GQMPRGLPGRGRPAGPAGARGNDGATGAAGPPTGAGPPGPPGAVGAGEAGP 355
QY 151 GPPGAGPAGRGRGQPP-----AGSPGPGGLPGPA 180
DB 356 GPRGSEGPQVGRGPPGPPGAGAGAGNPGADQPGKANGANGAPGFGARGPS 415
QY 181 GPPGAGKPGQGVFGDLGAP-----GPSGA-----GEPGPT 213
DB 416 GPQGGPPGPKNGSGEFCAGSGKDTGAKGEPVGVQPPGAGEGKRGARFPT 475
QY 214 GLPFPGRGSGRFPFGADGAGVAGPKPAGERSGPGAGKSGPAGRPGAGLPKAK 273
DB 476 GLPFPGRGSGRFPFGADGAGVAGPKPAGERSGPGAGKSGFAGRPGAGLPKAK 535
QY 274 GLTSGPSGPDGKTGPPGAGODRPPGPPGARGOAGVMGPPKGAAGEPKAGER 333
DB 536 GLTSGPSGPDGKTGPPGAGODRPPGPPGARGOAGVMGPPKGAAGEPKAGER 595
QY 334 GVPGPPGAVGAGKDGAGAGCPGPPGAPGAGEGQPGAGSPGQGLPAGPPGEAGKP 393
DB 596 GVFPGPPGAVGAGKDGAGAGCPGPPGAPGAGEGQPGAGSPGQGLPAGPPGEAGKP 655
QY 394 GEQGVPGDLGAPGSGPAGE-----PGP-----TGLGPP 423
DB 656 GEQGVPGDLGAPGSGPARGERFPGERGVQPPGPPGPRGANGAPNDGAKDAGAPCAP 715
QY 424 GERGGPSGRFPAGDAGVAGPKPAGERSGPGAGKSGPGE-----AGRPGAGLP 474
DB 716 GSQAPGLQMGGERGAGLPKPKDRDAGPKNDGSGFKDGVRLTGLTPIGPPGAPAG 775
QY 475 GAK-----GLTSGPSGPDGKTGPPGAGODRPPGPPGARGO----- 515
DB 776 GDKSGSPGAPGTGARGAPDRGEPGPPGAPGAPGADGPPGAKGEPCDAGAKDA 835
QY 516 -----AGVMGFPKGAAGEPKAGERGVFPFPGA-----VGPAGKDGAGAGQPP 561

; Sequence 1, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type I
US-08-931-820-1

Query Match 59.2%; Score 2746; DB 3; Length 1057;
Best Local Similarity 56.3%; Pred. No. 3.7e-170;
Matches 556; Conservative 32; Mismatches 232; Indels 168; Gaps 18;

QY 1 GPPGEP-----GPTGLPFPGRGSGRFGADGAGVAGKGPAGERSP----- 45
DB 42 GPRGEPGPGASGPMGPRGPPGPKNGDDGAGKPRFRGERGPPGPGQARGLPGTAGLP 101
QY 46 -----GPAGPK-----GSPGAGRPGAGLPKAKGLTSGPSGPG----- 80
DB 102 GMKXGRHFGSLDGAKDAGAPKPGFSGFGENGAFQWGRGLPGERGRPGAGPAGAR 161
QY 81 --DGKT---GPPGAGODRPPGPPGARGO-----GVMGFPKGAAGEP 123
DB 162 GNDGATGAAGPPGPTGAGPPGPPGAVGAKGAGPQGRGSEGPQGVRRGPPGPGAGAA 221

; Sequence 1, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type I
US-08-931-820-1

Query Match 59.2%; Score 2746; DB 3; Length 1057;
Best Local Similarity 56.3%; Pred. No. 3.7e-170;
Matches 556; Conservative 32; Mismatches 232; Indels 168; Gaps 18;

QY 1 GPPGEP-----GPTGLPFPGRGSGRFGADGAGVAGKGPAGERSP----- 45
DB 42 GPRGEPGPGASGPMGPRGPPGPKNGDDGAGKPRFRGERGPPGPGQARGLPGTAGLP 101
QY 46 -----GPAGPK-----GSPGAGRPGAGLPKAKGLTSGPSGPG----- 80
DB 102 GMKXGRHFGSLDGAKDAGAPKPGFSGFGENGAFQWGRGLPGERGRPGAGPAGAR 161
QY 81 --DGKT---GPPGAGODRPPGPPGARGO-----GVMGFPKGAAGEP 123
DB 162 GNDGATGAAGPPGPTGAGPPGPPGAVGAKGAGPQGRGSEGPQGVRRGPPGPGAGAA 221
```

Qy	124	GKAGERGVP	PPCAVGP	PAKGCB	EAGACQ	PPGAPG	AGERGE	QCPAGS	PFQGL	PGAPGP	183																			
Db	222	GPAGNPG	DADQ	QBCAK	ANGAPCI	AGAPFP	PCARPS	QPGFG	FTGARG	LIVSEP	PGAGSK	281																		
Qy	184	GKAGKPG	EGGV	PCDIL	CAPPS	-----	GPAGE	PPTGL	PPGPP	GERG	PGSR	FFPCAD	237																	
Db	282	GESGNK	GEPS	AGSP	PGPPG	PSGEG	KGRN	CEAGS	ACFP	PPGL	RGPP	RSRFF	PCAD	341																
Qy	238	GPKPAGER	GSPCP	AGPK	SGSP	CEAGR	PCEAGL	PKAGL	TGSP	PGSP	PGD	DKT	GTP	PGAG	297															
Db	342	GPKPAGER	SGSP	AGPK	SGSP	CEAGR	PCEAGL	PKAGL	TGSP	PGSP	PGD	DKT	GTP	PGAG	401															
Qy	298	GRGPPGP	PGARG	QAGV	NGFP	PKPKA	GERG	PKK	AGER	VP	PCAV	GP	PACK	DEAG	QAQPP	357														
Db	402	GRGPPGP	PGARG	QAGV	NGFP	PKPKA	GERG	PKK	AGER	VP	PCAV	GP	PACK	DEAG	QAQPP	461														
Qy	358	GPAGPAGER	GEO	PPAGS	PGF	QGL	PCAP	PPG	EA	CKP	GE	QV	PP	DL	CGAP	SPGAGE	--- 413													
Db	462	GPAGPAGER	GEO	PPAGS	PGF	QGL	PCAP	PPG	EA	CKP	GE	QV	PP	DL	CGAP	SPGAGE	--- 521													
Qy	414	-----	PGP	-----	-----	-----	TGL	CP	GER	G	RSR	FF	PCAD	G	VAG	PKPA	447													
Db	522	GERGVQ	GGPP	GPAG	PR	ANGA	PCND	AK	GD	AGAP	GS	QAGL	QMP	GER	GA	GLPK	581													
Qy	448	GERGSF	GPAG	PK	SGS	PE	-----	AGR	P	CE	AGL	PK	AK	-----	GLT	GS	PS	GP	D	489										
Db	582	GDRG	DAG	PK	AD	SG	PKD	VG	R	L	T	G	P	P	G	P	AG	P	D	G	641									
Qy	490	GKTG	PP	PAG	QD	R	GP	PP	G	ARG	Q	-----	AGV	M	FP	PF	PK	AA	GE	PKA	534									
Db	642	GF	GG	PP	GPAG	FAG	PP	GD	AG	Q	G	AK	GE	PD	AG	AK	GD	AG	PP	GP	AG	701								
Qy	535	GERGV	PG	PPGA	-----	-----	VGP	AK	D	GE	AG	AC	Q	PP	GP	PAGP	---	AGER	GE	Q	CP	AGS	PF	Q	582					
Db	702	GARG	SAG	PP	GAT	PP	GA	GR	V	GP	PS	GN	AG	PP	GP	CP	AK	GE	G	K	GR	GE	T	CP	AGS	PE	GE	761		
Qy	583	GL	PPAG	PP	GE	AK	GE	Q	V	PD	L	AG	PP	GS	GP	AGE	PG	PT	GL	PP	GP	GER	GG	PS	RR	FP	GD	642		
Db	762	GP	PP	PP	GPAGE	KS	GD	AG	P	AG	P	T	GP	OG	I	AG	OR	V	VL	PP	QR	GER	FF	PL	LP	SG	SEP	821		
Qy	643	GVAG	PK	PAG	BER	GS	PG	PAGP	---	KG	S	P	CE	AG	R	P	CE	AGL	PK	AK	L	T	GS	PS	GP	DP	DKT	699		
Db	822	GKQ	GPS	GAS	GER	PP	GP	NG	PP	GL	AG	PP	GE	S	GR	EG	A	PCA	EGS	PCR	D	GS	PCA	K	DR	GE	T	CPA	881	
Qy	700	GPAG	QD	GR	PP	PP	GPARG	QAGV	NG	FF	PKPKA	GERG	PKK	AGER	VP	PCAV	GP	PACK	DEAG	QAQ	GE	---	758							
Db	882	GP	PC	FAG	PAG	PP	VP	PA	GS	G	D	R	GE	T	GP	AG	P	3	VP	CA	GER	PAG	PP	GP	R	DK	GE	T	GEQ	941
Qy	759	-----	-----	AGA	Q	RR	PP	-----	-----	-----	AGP	A	GER	GE	Q	CP	AGS	---	GFQ	GL	LP	CPA	792							
Db	942	GDRG	K	IGH	R	SG	L	Q	CP	PP	PGS	P	GE	Q	PS	GA	GS	AG	P	GP	PGS	AG	A	PK	G	DL	N	GL	DP	1001
Qy	793	GP	PE	BAG	K	PG	BQ	V	PD	L	GA	P	PS	GP	SG	AP	820													
Db	1002	GP	PP	GR	GR	T	G	DAG	P	V	GP	PP	GP	PP	GP	PP	GP	1029												

RESULT 5

US-08-963-825-18

US-08-583-825-18
; Sequence 18, Application US/08963825

; Patent No. 6110689

; GENERAL INFORMATION:

APPLICANT: Qvist, Per

APPLICANT: Bonde, Martin

; TITLE OF INVENTION: A Method for Assaying Collagen Fragments

; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the

TITLE OF INVENTION: Method and Use of the Method to Diagnose the

; TITLE OF INVENTION: DIS

; NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:

; CORRESPONDENCE ADDRESS:
; ADDRESS: Derby & Derby DC

; ADDRESS: Darby & Darby
; STREET: 805 Third Avenue

CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/963,825

Query Match 58.3%; Score 2704; DB 3; Length 1341;
Best Local Similarity 57.3%; Pred. No. 2.3e-167;
Matches 542; Conservative 28; Mismatches 244; Indels 132

[illegible]

Db 601 GLQMPGEGAGLPGPKGDRGADGPKGADGSPKGVGLTGPIGPPGAGAPGDKGES 660
Qy 478 -----GLTSPGSGPDGKTPGPPGAGQDGRGPPGPPGARGQ-----A 516
Db 661 GPSGPAFTGARGAPGDRGPPGPPGAGFAGPPGADGQPKAGGEPDAGAKGAGPPGPA 720
Qy 517 GVMGFPKGAAFGPKAGRGVPPGPA-----VGPAGKDGAGAGQPPGAGP- 566
Db 721 GPAGFPPIGNVGAFAKAGARGAGFPAGTCFPGAAAGRVGPPGSGNAGPPGPPGAGKE 780
Qy 567 --AGERGQGPAGSGFQGLPLGPPAGPGEAGKPGQGVGDLGAPGSGPAGSGFTGLP 624
Db 781 GKGPRGTGTAGRGVGVPPGPPGAGKSGSPGADGAGAPGCTPPGQAGQGVVGLP 840
Qy 625 GPPGGRGSGRFPFGADGAVAGPKGAGRGSGPPAGP---KSGSPGAGRPGEAGLPGAK 681
Db 841 GQRGEGFPGLPGSGFPKGKSGASGERGPPGMPGPPGLAGPPGSGREGAPGAGSP 900
Qy 682 GLTSPGSGPDGKTPGPPGAGQDGRFPGPPGARGQAGVMPGPPKGAAGFPKGAGER 741
Db 901 GRDGSPGAKGDRGTGTPAGPPGAGXGAGAPGVGPAKSGDRGTGTGTPAGPAGVPVAGAR 960
Qy 742 GVPGPPGAVGPAKDG-----AGAGPPGP-----AGPAGERGEQ 777
Db 961 GPAGPQGRGDKGTGEGQDRGIKGRGFSGLQGGPPGPPGSGQSGASGAGPGRGPP 1020
Qy 778 GPAGSP---GPGGLPGAGPGEAGKPGEQGVGDLGAPGSPGAG 820
Db 1021 GSAGAPKGLGGLPIGPPGPRGRTCDAGVPVPPGPPGPPG 1066

RESULT 6
US-09-500-811-18
; Sequence 18, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Quist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
; US-09-500-811-18

Query Match 58.3%; Score 2704; DB 3; Length 1341;
Best Local Similarity 57.3%; Pred. No. 2.3e-167;
Matches 542; Conservative 28; Mismatches 244; Indels 132; Gaps 15;

Qy 1 GPPGE---PGPTGLPGPPGGERGPPGSRGF-----PCADGVAGPKGAPAGER 42
Db 127 GPPGRDGIPOQGLPGPPGPPGPPGPPGLGGNFAPQLYCYDEKSTGTSVPGPMGSPGR 186
Qy 43 GSPGAPAGPKSGSAGEAGRPGEAGLPGAKGLTSGPSGPDGKTPGPPGAGQDGRP---GPP 99
Db 187 GLGPPGAGPZGFZGPPGZPGASGPMGPPGPPGPPGPKGBGBGZAGKPRGPRGZRP 246
Qy 100 GPPGARGQAGVMGFPKPGKGAAGEPKGAGERGVPPGAVGVPAGKDGEGAGAOQPPGAPPA 159
Db 247 GPZGARGLPCTAGLPQMKHRCFSLBGAKGABAGAPGKZPSGSGZBGAFCZMGPPGPK 306
Qy 160 GERGQGPAGSPGFOGLPGPAGPGEAGKPGQGVGDLGAPGSGPAGEPGPTGLP 219
Db 307 GNSGE-----PCAPGSKGDTCAKGEPPGVQVQPPGAPGEGKRGARGEPGPTGLP 360
Qy 220 GSRGGPSRSGFPGADGVAGPKGAPGERSGPPGAPKSGSFGGAGRPGEAGLPGAKGLTSP 279
Db 361 GERGGPSRSGFPGADGVAGPKGAPGERSGPPGAPKSGSFGGAGRPGEAGLPGAKGLTSP 420
Qy 280 GSPGPDGKTPGPPGAGQDGRPFPGARGQAGVMGFPKPGKGAAGEPKGAGERGVPPGP 339
Db 421 GSPGPDGKTPGPPGAGQDGRPFPGARGQAGVMGFPKPGKGAAGEPKKAGERGVPPGP 480
Qy 340 GAVGPAKDGEGAGAOQPPGPPGAPGAGEGEOGAGSGPQGLPGPAGPGEAGKPGEQVP 399
Db 481 GAVGPAKDGEGAGAOQPPGPPGAPGAGEGEOGAGSGPQGLPGPAGPGEAGKPGEQVP 540
Qy 400 GDLGAPGSGPAGE-----PGR-----TGLPDPGPPGERGPP 429
Db 541 GDLGAPGSGPAGE-----PGR-----TGLPDPGPPGERGPP 600
Qy 430 GSRGFPAGDGVAGPKGAPGERSGPPGAPKSGSFG-----AGRPEAGLPGAK---- 477
Db 601 GLQMPGEGAGLPGPKGDRGADGPKGADGSGKDGVRGLTGPIGPPGAGAPGDKGES 660
Qy 478 -----GLTSPGSGPDGKTPGPPGAGQDGRGPPGPPGARGQ-----A 516
Db 661 GSPGPAFTGARGAPGDRGEPGPPGPPGAGPAGPPGADGQPKAGGEPDAGAKGAGPPGPA 720
Qy 517 GVMGFPKGAAAGEPKKAGERGVPPGPPGA-----VGPAGKDGEGAGAOQPPGAPG- 566
Db 721 GPAGFPPIGNVGAFAKAGARGAGFPAGTCFPGAAAGRVGPPGSGNAGPPGPPGAGKE 780
Qy 567 --AGERGQGPAGSGFQGLPGPAGPGEAGKPGQGVGDLGAPGSGPAGSGFTGLP 624
Db 781 GKGPRGTGTAGRGVGVPPGPPGAGKSGSPGADGAGAPGCTPPGQAGQGVVGLP 840
Qy 625 GPPGGRGSGRFPFGADGAVAGPKGAPGERSGPPAGP---KSGSPGAGRPGEAGLPGAK 681
Db 841 GQRGEGFPGLPGSGEPKQKSGASGERGPPGMPGPPGLAGPPGSGREGAPGAGSP 900
Qy 682 GLTSPGSGPDGKTPGPPGAGQDGRFPGPPGARGQAGVMPGPPKGAAGFPKGAGER 741
Db 901 GRDGSPGAKGDRGTGTGAPGPPGAGXGAGAPGVGPAKSGDRGTGTGTPAGPAGVPVAGAR 960
Qy 742 GVPGPPGAVGPAKDG-----AGAGPPGP-----AGPAGERGEQ 777
Db 961 GPAGPQGRGDKGTGEGQDRGIKGRGFSGLQGGPPGPPGSGQSGASGAGPGRGPP 1020


```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type II
; US-08-931-820-3

Query Match          53.4%; Score 2476; DB 3; Length 1060;
Best Local Similarity 50.9%; Pred. No. 1e-152;
Matches 503; Conservative 45; Mismatches 272; Indels 168; Gaps 17;

QY 1 GPPGPGTGLPVPGRGSGRFPDADGVAGPKGPAGERGSPGPAKGSFGEAGRP 60
DB 44 GNPGEPGPGVSGPMRPPGPPKPGDDGEAGKAGKAGRGPPGQARGPFTGGLP 103
QY 61 -----GEAGLPCAKLTCSP---GSPG-----DGKTGPPGPA--- 90
DB 104 GVKHGRVPLDGAKEAGAPGVKGESGPGNPGMGRGLPGRGRTGPAGAAGAR 163
QY 91 QDGRPGPPGPPGARGQAGVMGFP-----GPKGAAGEPPKAGRGVPPGPAVGPA 141
DB 164 GNDGQPGAGPPGVPVAGGPGFPGAPCAKCEAGTGAEGEAGRGPRGPTGSPGPA 223
QY 142 -----GKDAGAGQPPGPAAGPAGERGEGQAPGPGFQGLPGPAGPPGAGRGPGGVP 195
DB 224 GASGNPPTGDPGAKSGAGAPGAGFPGRGPPDPQATGFLGPKGTGTGPIAGFK 283
QY 196 GD---LGAPGSGPAGPPGTLPL-----GPPRGGPGSGRPPGADGVA 237
DB 284 GEQPKGEFPAGPQAGPAGGEGKRGARCEPGCVGPIGPPGERGAPGNRGPPQDGLA 343
QY 238 GPKGPAERGPAGPAGPKGSGEAGRGEBAGLPAKGLTSGPSGPDGKTGPPGAGQD 297
DB 344 GPKGAPGRGPGSLAGPKGANGDPERPCEPGLPGARGLTGRPDAGPQGVGSPGAPGD 403
QY 298 GRPGPPGPPGARGAGVMPGPKGAAGEPKKAGRGVPPGPAVAGKDGAGAGGPP 357
DB 404 GRPDPGQAGRGQGVNMGFPFGKANGEPKAGKGLPGAPLRLPGKDGETGABGPP 453
QY 358 GPAGPAGRGQGPAGSPGFGGLPAGPAGPGEAGKPGEGVPGDLGAPGSPGAGE--- 413
DB 464 GPAGPAGRGQGAPGSPGFGGLPAGPAGPGEAGKPGDQGVPEAGAPGLVGRGERGP 523
QY 414 -----PGTGLP-----GPPRGGPGSRGPPGADGVAGKGP 447
DB 524 GERGSPGAQGLQGRGLPFGTPTDTPKAGSAPGAPGPPQAQPPGLQGMPPGERGAAGTGP 583
QY 448 GERGS-----PAPAGKSPGEGAGRPGEAGLPAGKAGTGTSP 483
DB 584 GDRGDVGEKPEGAPKDGGRGLTGP:GPPCPAGANGKEGVEGPPGPPAGSAGARGAPGER 643
QY 484 GSPGPDGKTGPPGAGQDGRGPPGPPGARGQAGVMGFPKPKAAGEP-----GKA 534
DB 644 GETGPTGTSIAGPFGADGQFGAKGEGQAGQKGDAGAPGQPGSPGAPGQGTGTGPK 703
QY 535 GERGVPPGPA-----VGPAGKDGAGAGAPPPGPA---GPAGRGQGPAGSPGFQ 582
DB 704 GARGAQGPPGATGPPGAAGRVGPPSGNPNPFPPPGPPGSKDGPKARGDSGPPGAGERP 763
QY 583 GLPFPAGPPGAGKPGGQGVPLDGLAGPSPGAPGEPFTGLPFPGERGSGRGPFGAD 642
DB 764 GLOQFAPPPGKGFPGDDGSGAEGPPGQGLAGQIRGIVLPGQGRGRRPGLPGPSGEP 823
```

```

QY 643 GVAGPKGPAGERSGPPAGPKGSPGEAGRPGEAGLPGA-----KGLT 684
DB 824 CQQGAPGASGDRGPPGPPGVLGTGPAGEPGRGSPGADGPPRDRGAAAGVKDRGETCAV 883
QY 685 GSPGSPGPPGKTGPPGP-----AGQDRPGRPPGARGQAGVMGFPKPKAAGEP 735
DB 884 GAPGAPGPPGSPGAPGTPKQGRGEAGAQGPMGSPGAGARGIQGPGPRGDKGEAGEP 943
QY 736 GK---AGERGVPPGPAVGPAGKDGEGAGAQGPPGPAERGERGEGQAGSPGFGQLPGPA 792
DB 944 GERGLKGRHGTGLQGLPGLPSPGSDQAGSAGPAGSGRPPGPPGVPSPGKDGANGIPGI 1003
QY 793 GPPGAGKPGEGQVPGDLGAPGPPSPAG 820
DB 1004 GPPGPRGRSGETGPAGPFGNPPGPPG 1031

RESULT 13
US-08-963-825-20
; Sequence 20, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963.825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187.319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogozis, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
; US-08-963-825-20

Query Match          53.4%; Score 2476; DB 3; Length 1418;
Best Local Similarity 50.9%; Pred. No. 1.3e-152;
Matches 503; Conservative 45; Mismatches 272; Indels 168; Gaps 17;

QY 1 GPPGPGTGLPVPGRGSGRFPDADGVAGPKGPAGERGSPGPAKGSFGEAGRP 60
```

Db	156	GNPGEPEPGVSMGMRGPPGPKGDDDEAGKPKAGERPPGPGQARGPPGTGCLP	215
Qy	61	-----GBAGLPGAKGLTGP-----GSPGF-----DGKTPPGPA---90	
Db	216	GVKHRGYPCLGDAKEAGAPGVKSGSGSPGNGSPGWMGRGLPGERGRTPAGAAGAR	275
Qy	91	QDGRPGPPPGARGQAQVWNGPP-----GPKGAAGEPKKAGRGVGPSPGAVGPA	141
Db	276	GNDQPGPAPFPVPVGPAGPGFPAGPAKXGEAGFTGARGPEGAQGRGPGFTGSPGA	335
Qy	142	-----CKDGEAGAQQSPGAPGAGERGEQGPAGSPQGLPGPAGPGGAGKPGEGQV	195
Db	336	GASGNPTDGI PKAKGSAGAPGAGPFGPRGPPDQATGPLGPKGQTGREGIAGFK	395
Qy	196	GD---LGAPSPSPAGBPGTGLP-----GPPRGGQPGSGRPPGADGVA	237
Db	396	GEQPKPEPGFAPQGAAGPAGBEGKRGARGPGGVPI GPPGSRGAPGNRPGPGDGLA	455
Qy	238	GPXGPACERGSPPGAPKPGSPGAGRPCEAGLPGAKGLTSPGSPGPDGTPPGPAGQD	297
Db	456	GPKGAPGERSPGLAGPKGANGDCPRGEPGLPGARGLTGPDAGPQGVKVPSGAFGED	515
Qy	298	GRPPPGPPGQAQQAQVWNGPPGPKGAAGEPKKAGRGVGPSPGAVGPAKDXGEAGAQQPP	357
Db	516	GRDGPFGFQARGQFQVWNGPPGPKGANGEPGKAGEKGLPGAPGLRGLPKDGETGAE	575
Qy	358	GPAGPAGERGEQGPAGSPGQGLPGPAGPGEACKPGBOQVPGDLGAPGSGPAGE---	413
Db	576	GPAGPAGERGEQGPAGSPGQGLPGPFPGEGGKPGDQGVFGBAGAPGLVGRKGRGFP	635
Qy	414	-----PGTGLP-----GPPGERGPGSRGPPGADGVAGPKGPA	447
Db	636	GERSGPCAQGLQGRGLPGTPTDGPKGASGPAGPPCAQCPFGQGMWGERGAAGIAGPK	695
Qy	448	GERGS-----TPGAPKSGSPGAGRPGGAGLPGAKGLTGP	483
Db	696	GDRGUVGEKPEGAPGKDXGRGLTGP GPGFAGANGKEKGVPPGPAGSAGARGPGER	755
Qy	484	GSPFDGKTGPPGAGQDGRPPGPPGARGQAQVWNGPPGPKGAAGPP-----GKA	534
Db	756	GETGPTGSLTAGPPGADGQFGAKGEOGKADAGAPGQPSGAPPGQPGTGVTPK	815
Qy	535	GERGVPPGPA-----VGPAGKDEAGAAQGPQPA---GPAGERGEOGPAGSPGFQ	582
Db	816	GARGAQPPGATGTFPFAAGRVGPPGSGNGNPPFPFPSPGKDPGKARGDSGPPRAGEP	875
Qy	583	GLPGPAPGPEAGKPGGQGVPGDLGAPGSPGAPGPTGLPGPPGRRGGSPGSGFPDAD	642
Db	876	GLQGPAPGPGKSGFEGDDGSPGABGPPCPQLAQQRGIVGLPQQRGKGFPLFPSPGEP	935
Qy	643	GVAGFKPAGERGSPGAPGPKSGSPGEGAPRPEAGLPGA-----KGLT	684
Db	936	GQOQAPGASGDRGPPGVGPPGLTGAPGEPGREGSPGADGPPGRDGAAGVKXDRGETAV	995
Qy	685	GSPGSPGDGKTGPPGP-----AQDGRPPGPPGARGQAQVWNGFPKGAAGEP	735
Db	996	GAPGAPGPGSPGAGPTGKQGDGRGKGAQQPMGSPGAGARGIQGPQGPGRDGKEAGEP	1055
Qy	736	GK---ABERGVPFGVAGVPAGKDKGEAGAAQPPGAPGABRGQGPAGSPGFQCLGPPA	792
Db	1056	GERGLKHRGFTGLQLGPPGPGPSGDQASGAPGSGPRGPPGVGSPGKDGANGIPGI	1115
Qy	793	GPPEAKKPGGQGVFDGLGATGPGSGPAG	820
Db	1116	GPDPGRGRSGETGPAGPFGNCPGPPPG	1143

RESULT 14

US-09-010-999-1

US-09-010-999-1
: Sequence 1. Application US/09010999

Sequence 1, Applicant
: Patent No. 6132976

; Patient NO. 6132976
: GENERAL INFORMATION:

Db 456 GPKGAPGERGSLAGPKGANGDPGRPCGGLPGARGLTGRPDAGPKGVKPSGAPGED 515
Qy 298 GRPGRGPPGARGQGVNGVFPKGAAGEPKKAGRGVPGPGVAGVAGKDGAGAGQPP 357
Db 516 GRPGRGPPGARGQGVNGVFPKGAAGEPKKAGRGVPGPGVAGVAGKDGAGAGQPP 575
Qy 358 GPAGPAGRGQGPAGSGPFGQLPGPAGPPGAGKPGEQVPGDGLGAPGSGPAGE---- 413
Db 576 GPAGPAGRGQGPAGSGPFGQLPGPAGPPGAGKPGEQVPGDGLGAPGSGPAGE---- 635
Qy 414 -----PGTGLP----- 695
Db 636 GERGSPGAGLQGRGLPGTGTDPGKASGAPGPPGAGQPPGGLQGMPPGSGAGAGPK 695
Qy 448 GERGS----- 755
Db 696 GDRGDVGEKPEGAPKDGGRGLTGPTGPPGAGANGKEKGEVGPFGPAGSAGAGAPGER 755
Qy 484 GSPGPDGKTPGPPAGQDGRGPPGPPGARGQAGVMPGPPGKGAAGEP-----GKA 534
Db 756 GETGPPGTSGIAGPPGADGQCAKGEQGEAGKGDAGAPGQPSGAPGQPGTGTGTPK 815
Qy 535 GERGVGPPGA-----VGPAGKGEAGAGPPGPA---GPAGRGQGPAGSGPQF 582
Db 816 GARGAQGPATGPPGAGRGVPPGSGNPNPPGPPGSKDKGPKGARGSGPPGAGEP 875
Qy 583 GLPGPAGPPGAGKPGEQVFGDLGAPGSPGAPGFTGLPGPPGRRGPPGSGPPGAD 642
Db 876 GLQGPAGPPGKPGEDDPSGAGSPGQGLAQGIVGLPGQGRGPPGLPGPSGP 935
Qy 643 GVAGPKGAPGRCSPGAPGKSGGAGRPCEAGLPGA-----KGLT 684
Db 936 GQGPAGSGDRGPPGPPGLTGAPGPPGREGSPGADGPPGSDGAGVKGDRGTGAV 995
Qy 685 GSPSGPDGKTPGPPG-----AQDGRPPGPPGARGQAGVMPGPPGKGAAGEP 735
Db 996 GAPGAPGPPGPPGAPGTPGKQDRGAGAGAGQAGPPGPPGPPGPPGPPGPPGPPG 1055
Qy 736 GK---AGRGVPPGPPGAPGKDGAGAGAGQAGPPGPPGPPGPPGPPGPPGPPGPPG 792
Db 1056 GERGLKGRGTGLQLGPPGPPGSDGAGSPGPPGPPGPPGPPGPPGPPGPPGPPG 1115
Qy 793 GPCEAGKPGQGVPGDLGAPGPPGPPG 820
Db 1116 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1143

RESULT 15

US-09-500-811-20

; Sequence 20, Application US/09500811

; Patent No. 6323114

; GENERAL INFORMATION:

; APPLICANT: Qvist, Per

; TITLE OF INVENTION: A Method for Assaying Collagen Fragments

; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the

; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

; TITLE OF INVENTION: Disorders Associated with the Metabolism of

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Darby & Darby PC

; STREET: 805 Third Avenue

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/500,811

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
; US-09-500-811-20

Query Match 53.4%; Score 2476; DB 3; Length 1418;

Best Local Similarity 50.9%; Pred. No. 1.3e-152;

Matches 503; Conservative 45; Mismatches 272; Indels 168; Gaps 17;

Qy 1 GPGRGPPGTGLPGPPGRRGSGRFPAGDVGAPKPGAGRGSPGAPGKSGSGEAGRP 60

Db 156 GNPGEFPGVSGMPPGPPGPPGPKGDDGAGKPGKAGRGPPGPKARGFPGTGLP 215

Qy 61 -----GEAGLFGAKGLTGP-----GSPG-----DGKTGPPGPA--- 90

Db 216 GVKHGRGYPLDGAKEAGAFGVKSGSPGSGSPGPMGRGLPGRGRTGPAGAAG 275

Qy 91 GQDGRPPGPPGARGQAGVMPG-----GPKGAGPPGKAGRGVPPGPPGAPGA 141

Db 276 GNDGQFPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 335

Qy 142 -----GKDGAGAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 195

Db 336 GASGNPCTDGIPLGAKGSAGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPP 395

Qy 196 GD---LGAPSGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 237

Db 396 GEQGPKEFPAGPQAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 455

Qy 238 GPKGPAGRGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 297

Db 456 GPKGAPRGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 515

Qy 298 GRPGRGPPGARGQAGVMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 357

Db 516 GRPGRGPPGARGQAGVMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 575

Qy 358 GPAGPAGRGQGPAGSGPFGQLPGPAGPPGAGKPGEQVPGDGLGAPGSGPAGE---- 413

Db 576 GPAGPAGRGQGPAGSGPFGQLPGPAGPPGAGKPGEQVPGDGLGAPGSGPAGE---- 635

Qy 414 -----PGTGLP----- 695

Db 636 GERGSPGAGLQGRGLPGTGTDPGKASGAPGPPGAGQPPGGLQGMPPGSGAG 695

Qy 448 GERGS----- 755

Db 696 GDRGDVGEKPEGAPKDGGRGLTGPTGPPGAGANGKEKGEVGPFGPAGSAG 755

Qy 484 GSPGPDGKTPGPPAGQDGRGPPGPPGARGQAGVMPGPPGKGAAGEP-----GKA 534

Db 756 GETGPPGTSGIAGPPGADGQCAKGEQGEAGKGDAGAPGQPSGAPGQPGTGTG 815

Search completed: October 18, 2004, 13:44:02
Job time : 39.5431 secs

This Page Blank (uspio)

Db 313 GRPFGPPAGARGNDGATGAAGPPGPTGAPGPPPGAVGAKGAGPQAGRGSEGPQGV 372
Qy 148 GAGQPPGPPAGPAGRGEGP-----AGSPGQGLPGPAGPPGEGAGKPEQ 192
Db 373 GEPFGPPGAGAGPAGNPGADQCGKAGANGAPGIAGAPFGGARGSGGPPGPPGPK 432
Qy 193 GVPGLGAP-----GPSGA-----GEPGPTGLPGPPGERGGP 225
Db 433 GNSGPPGAPGSKGDTGAKGEPPTGVQPPGAGEGKRGARGGPPAGLPGPPGERGGP 492
Qy 226 GSRGPPGADGVAGPKPAGRGSGPAGPKSPGEGAGPGEAGLPGAKGLTSGPSGPPD 285
Db 493 GSRGPPGADGVAGPKPAGRGSGPAGPKSPGEGAGPGEAGLPGAKGLTSGPSGPPD 552
Qy 286 GKTGPPGAGQDGRPPGPPGARGQAGVMGPPGPKAAAGBPAGKAGRGVPPGPPGAVGA 345
Db 553 GKTGPPGAGQDGRPPGPPGARGQAGVMGPPGPKAAAGBPAGKAGRGVPPGPPGAVGA 612
Qy 346 KDGEGAGQPPGAGPAGRGEGPAGSPGQGLPGPAGPPGAGKPGGQGVPGDLGAP 405
Db 613 KDGEGAGQPPGAGPAGRGEGPAGSPGQGLPGPAGPPGAGKPGGQGVPGDLGAP 672
Qy 406 GPSGPAGE-----PGP-----TGLPGERGPPGSRGPP 435
Db 673 GPSGARGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 732
Qy 436 GADGAGPKGPPAGRGSGPPAGPKSPG-----AGRPGEAGLPGAKGLTSGPS 486
Db 733 GERGAAGLPGKDRGDAGPKGADGAPCKDGVRLTGPPIGPPGAGAPGDKGTPSGPA 792
Qy 487 GPDGKTGPPGAGQDGRPP-----PGPPGARGQAGVMGF-----GSPGPD---GKTGPPG 701
Db 973 SGPSGERGPPGMPG---PGLAGPPBSGREGAPGAGSGPDRDAGAPCKDREGSGPAG 1029
Qy 702 AGQDGRPPGPPGARGQAGVMGPPGPKAAAGBPAGKAGRGVPPGPPGAVGAPKAGDE--- 758
Db 1030 PGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1089
Qy 759 -----ACAQPPGP-----AGPAGRGEGPAGSP-----GFGGLPGPAG 794
Db 1090 RGIKHGRGSLQPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1149
Qy 795 PGEAGKPGEGVPGDLGAPGSPGAG 820
Db 1150 PGPGRGTGAGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1175

RESULT 2

US-10-402-072A-8

; Sequence 8, Application US/10402072A

; Publication No. US20040018592A1

; GENERAL INFORMATION:

; APPLICANT: Bell, Marcum P.

; APPLICANT: Neff, Thomas B.

; APPLICANT: Polarek, James W.

; APPLICANT: Slesley, Todd W.

; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS

; FILE REFERENCE: FPO402.2 CON

; CURRENT FILING DATE: 2003-03-26

; PRIOR APPLICATION NUMBER: US 09/709,700

; PRIOR FILING DATE: 2000-11-10

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 8

; LENGTH: 1449

; TYPE: PRT

; ORGANISM: Sus scrofa

US-10-402-072A-8

Query Match 59.6%; Score 2765.5; DB 15; Length 1449;

Best Local Similarity 56.7%; Pred. No. 4.5e-131; Indels 169; Gaps 21;

Matches 559; Conservative 32; Mismatches 226

Qy 1 GPPGPGPTGLPGPPGERGGPQS-----RGFFGA-----DGVAGPKPAGRGSGPQA 48
Db 193 GPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 252
Qy 49 GPKGSPGAGRP-----CEAGLPGAKGLTSGPSGPPGPKTGPAGPAGQD 93
Db 253 GARGLPGTAGLPGMKHGRGFSGLDGAAGPAGPKGPPGPPGPPGPPGPPGPPGPP 312
Qy 94 GRPFGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 147
Db 313 GRPFGPPGAGARGNDGATGAAGPPGPTGAPGPPGAVGAKGAGPQAGRGSEGPQGV 372
Qy 148 GAGQPPGPPAGPAGRGEGP-----AGSPGQGLPGPAGPPGEGAGKPEQ 192
Db 373 GEPFGPPGAGAGPAGNPGADQCGKAGANGAPGIAGAPFGGARGSGGPPGPPGPK 432
Qy 193 GVPGLGAP-----GPSGA-----GEPGPTGLPGPPGERGGP 225
Db 433 GNSGPPGAPGSKGDTGAKGEPPTGVQPPGAGEGKRGARGGPPAGLPGPPGERGGP 492
Qy 226 GSRGPPGADGVAGPKPAGRGSGPAGPKSPGEGAGPGEAGLPGAKGLTSGPSGPPD 285
Db 493 GSRGPPGADGVAGPKPAGRGSGPAGPKSPGEGAGPGEAGLPGAKGLTSGPSGPPD 552
Qy 286 GKTGPPGAGQDGRPPGPPGARGQAGVMGPPGPKAAAGBPAGKAGRGVPPGPPGAVGA 345
Db 553 GKTGPPGAGQDGRPPGPPGARGQAGVMGPPGPKAAAGBPAGKAGRGVPPGPPGAVGA 612
Qy 346 KDGEGAGQPPGAGPAGRGEGPAGSPGQGLPGPAGPPGAGKPGGQGVPGDLGAP 405
Db 613 KDGEGAGQPPGAGPAGRGEGPAGSPGQGLPGPAGPPGAGKPGGQGVPGDLGAP 672
Qy 406 GPSGPAGE-----PGP-----TGLPGERGPPGSRGPP 435
Db 673 GPSGARGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 732
Qy 436 GADGAGPKGPPAGRGSGPPAGPKSPG-----AGRPGEAGLPGAKGLTSGPS 486
Db 733 GERGAAGLPGKDRGDAGPKGADGAPCKDGVRLTGPPIGPPGAGAPGDKGTPSGPA 792
Qy 487 GPDGKTGPPGAGQDGRPP-----PGPPGARGQAGVMGF-----GSPGPD---GKTGPPG 539
Db 973 SGPSGERGPPGMPG---PGLAGPPBSGREGAPGAGSGPDRDAGAPCKDREGSGPAG 852
Qy 540 PGPPCA-----VGPAGKDGAGAGQPPGPPGAPAGE-----RGEQPPAGSPGQGLPGP 587
Db 853 AGPPGATGPPGAAAGRVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 912
Qy 588 AGPPGAGKPGEGVPGDLGAPGSPGAG-----GSPGPD---GKTGPPG 701
Db 913 GPPAGEKSPGADGAPAGPTGPGQIAGQGVVGLFQGRGERGPPGLPGPSGPPGQGP 792
Qy 648 KGPAGERSGPPAGPKSPGAGRGPPGAGLPGAKGLTSGP-----GSPGPD---GKTGPPG 701
Db 973 SGPSGERGPPGMPG---PGLAGPPBSGREGAPGAGSGPDRDAGAPCKDREGSGPAG 1029
Qy 702 AGQDGRPPGPPGARGQAGVMGPPGPKAAAGBPAGKAGRGVPPGPPGAVGAPKAGDE--- 758
Db 1030 PGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1089
Qy 759 -----ACAQPPGP-----AGPAGRGEGPAGSP-----GFGGLPGPAG 794
Db 1090 RGIKHGRGSLQPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1149
Qy 795 PGEAGKPGEGVPGDLGAPGSPGAG 820
Db 1150 PGPGRGTGAGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1175

Query Match 59.6%; Score 2765; DB 15; Length 1107;
Best Local Similarity 55.4%; Pred. No. 3.9e-131;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGPGTGLPGERGSGRFPAGDVGAGPKGAGRGSPGAGPKGSGEAG--- 57
DB 18 GPMGSGRGLPFGAGPQGFQPPGEPGEGASGPMGPRGPPGKNGDDGEAGKP 77
QY 58 GRPGAGLP---GAKGLTSGSGSPG-----DGKTPPGAGQDGRPPGPPGAR 105
DB 78 GRGGRGPPGQAGRLGTAGLPGMKHGRFSLDGAKGDAGPAGPKGSGSGENGAP 137
QY 106 GQAGVMPGPGKGAAGEPKAGER-----GVPPGAGVGA-----GKDGAGAQ 150
DB 138 GQMGPRGLPGRGRGAGAPAGANDGATGAAGPPGTGTPAGPPGAGVAGKEAGPQ 197
QY 151 GPPGAGPAGRGEGQP-----AGSPGQGLPGPA 180
DB 198 GRGSGPGVGRGPPGAGAGPAGNPGADGQPKAGKANGAPGAGPAGPARGPS 257
QY 181 GPPGAGKGGVPGDLAGP-----GPSGA-----GEPGPT 213
DB 258 GQGGGPGPPKNGSGEPGAGSKGDTGAKGEPGVQGGPPGAGBEGKRGAGEPGPT 317
QY 214 GLPFGPPGRGSGRFPAGDVGAGPKGAGERSGPGAGPKGSGPAGRPGEGAGLPKAK 273
DB 318 GLPFGPPGRGSGRFPAGDVGAGPKGAGERSGPGAGPKGSGPAGRPGEGAGLPKAK 377
QY 274 GLTSGSGSPDGKTPPGPAGQDGRPPPPGARGQAGVMPGPKGAAGEPKAGER 333
DB 378 GLTSGSGSPDGKTPPGPAGQDGRPPPPGARGQAGVMPGPKGAAGEPKAGER 437
QY 334 GVPGPPGAVGAGKDGAGAGPAGPAGGERGQAGSPGFGQLPGPAGPGEAGKP 393
DB 438 GVPGPPGAVGAGKDGAGAGPAGPAGGERGQAGSPGFGQLPGPAGPGEAGKP 497
QY 394 GEQGVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 423
DB 498 GEQGVPGDLGAPGSGGARGGFPGERGVGPPGPPGAGPRGANGAPGNDGAKGDAGAPGAP 557
QY 424 GERGPGSGRPPGADVGAGPKGAGERSGPGAGPKGSGP-----AGRPGAGLP 474
DB 558 GSQAGPLQGMFGGKAGAGLPKPKDRDAGPKADSGFGDGVRLGTGPIGPPGAPGAP 617
QY 475 GAK-----GLTSGSGSPDGKTPPGPAGQDGRPPGPPGARGQ----- 515

DB 618 GDKGSGSPGAPGTGARGAPGDRGEPGPPGAGPAGPAGDQGFAGKGEPCDAGAKGDA 677
QY 516 -----AGVMPGPKGAAGEPKAGRGVPGPPGA-----VGPAGKGEAGAQGPP 561
DB 678 GPPGAPGAPGPPGPIGNVAGAPKAGARGSPGATGFPGAAGRVPGPSPGSGNAGPPGPP 737
QY 562 GRAGP---AGERGEGPAGSPGFGQLPGPAGPAGKRGSGVPGDLGAPGSPGAGEP 618
DB 738 GPAGKEGGKGRGETGPAGRGVEVPPGPPGAGKSGPAGDGPAGAGTTPGQIACQR 797
QY 619 GPTGLPFPGERGSGRFPAGDVGAGPKGAGERSGPPGAPG---XGSGEAGRPGEA 675
DB 798 GVVGLPGORGERGFPGLPGSGEPCKQKQPSGASGERGPPGMPGGLAGPPGSGREGAP 857
QY 676 GLPGAAGLTSGSGSPGDPKTPGPPAGQDGRPPGPPGARGQAGVMPGPKGAAGEP 735
DB 858 GAGSGPGRDGSFCAKGRGETGPAGPPGAPGAPGPPVGPAGKSGDRGETGPAGPAGPV 917
QY 736 GKAGERGVPPGAGVAGPKAGE-----AGAGGPPGP-----AGPA 771
DB 918 GPAGARGPAGPQGRDKGETGEGQDRGIKGRHGFSLGQPPGPGSPGEGQPSGASGPA 977
QY 772 GERGEGPAGSP---GFQGLPGPAGPAGKRGVPGDLGAPGSGPAG 820
DB 978 GPRGPPGSAGAPGKDLNGLPGFPGPPRGTGAGVMPGPPGPPGPPG 1029

RESULT 5

US-10-104-889-8
; Sequence 8, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: BRUSKIN, ELLIOT A.
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EASLE CVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1171 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-104-889-8

Db 318 GLPFPGERGGSGFTGADGVAGPKDPAAGERSGPGPAGKSGSGEAGRPGEAGLPGAK 377
Qy 274 GLTGSFGSPGDPKGTGPPGAGQDGRPGPPPGARQAGVGMFFPGKGAAGEPGKAGER 333
Db 378 GLTGSFGSPGDPKGTGPPGAGQDGRPGPPPGARQAGVGMFFPGKGAAGEPGKAGER 437
Qy 334 GVPFPPGAVGPKDGEAGAGQPPGAPGAGERGGGAGSGPGQGLPGRAGPGEAGKP 393
Db 438 GVPFPPGAVGPKDGEAGAGQPPGAPGAGERGGGAGSGPGQGLPGRAGPGEAGKP 497
Qy 394 GEQGVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 423
Db 498 GEQGVPGDLGAPGSGARGRGFFGPGVQPPGAPGPRGANGAPGNDGAKGDAGAPGAP 557
Qy 424 GERGGSGRFFGADGVAGPKDPAAGERSGPGPAGKSGSGE-----AGRPGEAGLP 474
Db 558 GSQGAPGLQMPGERGAAGLPKPKDGRDAGPKGADGSGKDGVRGLTGTPTGPPGAPAG 617
Qy 475 GAK-----GLTGSFGSPGDPKGTGPPGAGQDGRPGPPGARGO-----515
Db 618 GDXGESGSGPAGPTGARGAPDGRGEPGPPGAPGAGPPGADGQPKAGEFGDAGAKGDA 677
Qy 516 -----AGVMPGPKGAAGEPKKAGERGVPPGPA-----VGPAGKDEAGAGQPP 561
Db 678 GPPGAPGAPGPPGPIGNVGAAGKARGSGAGPPGATGFGAAGRVGPPGSGNAGPPGPP 737
Qy 562 GPAGP---AGERGEGPAGSGFQGLPGPAGPPGEGAGKPGSQGVPGDLGAPGSGPAGER 618
Db 738 GPAGKEGKGRGETGAPRGEVGPVPPGPPGAGKSGPGADGAPAGETGPGQIAGQR 797
Qy 619 GTGLPGERGGSGRFFGADGVAGPKDPAAGERSGPGPAGP---XGSGEAGRPGEA 675
Db 798 GVVGLPQGRGERFFGLPSPGCEPKQKPSGASGERGPPGMPGGLAGPPGSGREGAP 857
Qy 676 GLPAGKGLTGSFGSPGDPKGTGPPGAGQDGRPGPPPGARQAGVGMFFPGKGAAGEP 735
Db 858 GAEGSPGSDGSGAKGDRGETGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 917
Qy 736 GXAGERVGPPGAVGAPGAKGDE-----AGAGQPPG-----AGPA 771
Db 918 GPAGARGPAGPGRGDKETGEQDRIKGRHFGSLGQPPGPGSGEGOGPSGASGPA 977
Qy 772 GEREGQAGASP---GFGQLPGPAGPGEAGKPGQGVPGDLGAPGSGPAG 820
Db 978 GPRGPPGSAAGPKDGLNGLPGIFGPPGPRGRTGDAVPVGPVPPGPPGPPG 1029

RESULT 7

US-10-216-705-21
; Sequence 21, Application US/10216705
; Publication No. US20030096973A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me
; TITLE OF INVENTION: Obtaining Such and Their Uses
; FILE REFERENCE: 1149-3 DIV
; CURRENT APPLICATION NUMBER: US/10/216,705
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 09/331,347
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-216-705-21

Query Match 59.6%; Score 2765; DB 14; Length 1464;
Best Local Similarity 55.4%; Pred. No. 4,8e-131;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

Qy 1 GPPGEPGPTGLPFPGERGGSGRFFGADGVAGPKDPAAGERSGPGPAGKSGSGEAGRPGEAGLPGAK 57

Db 179 GPMVSGPRLGLPFPGAPGAPGQFQFPPEPGEAGSGPMGPRGPPGPKGNKDDGEACKP 238
Qy 58 GRPGEAGLP---GAKGLTGSFGSPGDP---DKGTGPPGAGQDGRPGPPPPGAR 105
Db 239 GRPGERGPPGQAGKGLPGLTAGLPGMKHGRGSLDGAKGADGAPGPKGEPGSGENGAP 298
Qy 106 QGAGVGMFFGPGKGAAGEPKKAGER-----GVPFPPGAVGPA-----GKDEAGAG 150
Db 299 GQNGPRLGFGERPGCAPGAPGARGNDGATGAAGPPGPTGPPGAPGFPFCAVAKGEAGPQ 358
Qy 151 GPPGAPGAPGAGEGEGP-----AGSPGFQGLPGPA 180
Db 359 GPRGSGPQVGEPCGPPGAPGAGNPGADGQPGAKGANGAPGIAGAPGFGARGPS 418
Qy 181 GPPGEAGKGBGQVPGDLGAP-----GPSGA-----GEPGT 213
Db 419 GPQPGGPPGPKNGSGEPGAPGSKDGTAKGPPGVGVQPPGPPAGEBKRGARGEPGT 478
Qy 214 GLPFPGERGGSGRFFGADGVAGPKDPAAGERSGPGPAGKSGSGEAGRPGEAGLPGAK 273
Db 479 GLPFPGERGGSGRFFGADGVAGPKDPAAGERSGPGPAGKSGSGEAGRPGEAGLPGAK 538
Qy 274 GLTGSFGSPGDPKGTGPPGAGQDGRPGPPPGARQAGVGMFFPGKGAAGEPGKAGER 333
Db 539 GLTGSFGSPGDPKGTGPPGAGQDGRPGPPPGARQAGVGMFFPGKGAAGEPGKAGER 598
Qy 334 GVPFPPGAVGPKDGEAGAGQPPGAPGAGERGGGAGSGPGQGLPGPAGPGEAGKP 393
Db 599 GVPFPPGAVGPKDGEAGAGQPPGAPGAGERGGGAGSGPGQGLPGPAGPGEAGKP 658
Qy 394 GEQGVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 423
Db 659 GEQGVPGDLGAPGSGARGRGFFGPGVQPPGAPGPRGANGAPGNDGAKGDAGAPGAP 718
Qy 424 GERGGSGRFFGADGVAGPKDPAAGERSGPGPAGKSGSGE-----AGRPGEAGLP 474
Db 719 GSQGAPGLQMPGERGAAGLPKPKDGRDAGPKGADGSGKDGVRGLTGTPTGPPGAPAG 778
Qy 475 GAK-----GLTGSFGSPGDPKGTGPPGAGQDGRPGPPGARGO-----515
Db 779 GDXGESGSGPAGPTGARGAPDGRGEPGPPGAPGAPGAPGAPGAPGAPGAPGAPGAP 838
Qy 516 -----AGVMPGPKGAAGEPKKAGERGVPPGPA-----VGPAGKDEAGAGQPP 561
Db 839 GPPGAPGAPGPPGPIGNVGAAGKARGSGAGPPGATGFGAAGRVGPPGSGNAGPPGPP 898
Qy 562 GPAGP---AGERGEGPAGSGFQGLPGPAGPGEAGKPGQGVPGDLGAPGSGPAGEP 618
Db 899 GPAGKEGKGRGETGAPRGEVGPVPPGPGAGEKSGFADGAPGAPGPTGPGQIAGQR 958
Qy 619 GPTGLPFPGERGGSGRFFGADGVAGPKDPAAGERSGPGPAGP---KSGPGEAGRPGEA 675
Db 959 GVVGLPQGRGERFFGLPSPGCEPKQKPSGASGERGPPGMPGGLAGPPGSGREGAP 1018
Qy 676 GLPAGKGLTGSFGSPGDPKGTGPPGAGQDGRPGPPPGARQAGVGMFFPGKGAAGEP 735
Db 1019 GAEGSPGSDGSGAKGDRGETGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 1078
Qy 736 GKAGERGVGPPGAVGAPGAKGDE-----AGAGQPPG-----AGPA 771
Db 1079 GPAGARGPAGPGRGDKETGEQDRIKGRHFGSLGQPPGPGSGEGOGPSGASGPA 1138
Qy 772 GEREGQAGASP---GFGQLPGPAGPGEAGKPGQGVPGDLGAPGSGPAG 820
Db 1139 GPRGPPGSAAGPKDGLNGLPGIFGPPGPRGRTGDAVPVGPVPPGPPGPPG 1190

RESULT 8

US-10-357-851-1
; Sequence 1, Application US/10357851
; Publication No. US20040151731A1
; GENERAL INFORMATION:

Db 659 GEQGVFCDLGAFCPSGARGGFGPGERGVQGGPAGPRGANGAPCNDGAKGDAGAPCAP 718
 QY 424 GERGGSGRFGGADGAGVAGKGPAGERSGPGAGKSGPGE-----AGRGEAGLP 474
 Db 719 GSGGAPLQCMFGERGAAGLPKGDGADGPKGADGSPGKGVGLTGTPIGPPGAPAG 778
 QY 475 GAK-----GLTSGPSGPDGKTGPPAGODGRPPGPPGARGQ-----515
 Db 779 GDKGESGSPGAPGTGARGAPGDRGEPGPPGAGFAGPGADGQSGAKGEPGDAGAKDA 838
 QY 516 -----AGVWFGPKGAGEPKGAGRGVPGPPGA-----VGPAGKGEAGAGQPP 561
 Db 839 GPPGAPGAPGPPPIGNVAPAKARGAGSGPAGTGFPGAAGRVGPPGPPSGNAGPPGP 898
 QY 562 GPAGP---AGERGEOGAGSPGFGGLPGPAGPPGAGKPGGQGVFCDLGAFCPSGAPAGEP 618
 Db 899 GPAGKEGGKPRGETGAPRGVEVGPAGGKSGADGAGAGTGTGPGIAGOR 958
 QY 619 GPTGLPGRGGRGSGRFGPCADGAGVAGKGPAGERSGPGAGP---KGSFGEAGRGEA 675
 Db 959 GVVGLPGORGERGFPGLFPGSGEPKQGPSCASGERGPPGPPGLAGPPGSGREGAP 1018
 QY 676 GLPGAKGLTSGPSGPDGKTGPPAGODGRPPGPPGARGQGVFCDLGAFCPSGAPAGEP 735
 Db 1019 GAEGSGRDSGFAKEDRGETGAPGAPCAPGAPGVPVAGKSGDRGETGAPGAPGV 1078
 QY 736 KAGBRGVGPPGAVGAPKXGDE-----AGAQGPPG-----AGPA 771
 Db 1079 GPAGARGPAGPQGPREDKGETGEQDRIKGRHFGSLQGGPPGSGEQQSGASGPA 1138
 QY 772 GERGEOGAGSP---GFOGLPGPAGPGEAGKPGEQVFDLGAFCPSGAPAG 820
 Db 1139 GPRGPPGAGAPGKDLNGLPGLPPIGPPRGTGDAGVGPVPPGPPGPPG 1190

RESULT 10
 US-10-788-792-150
 ; Sequence 150, Application US/10788792
 ; Publication No. US20040191819A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bayer Pharmaceuticals Corporation
 ; APPLICANT: Eveleigh, Deepa
 ; APPLICANT: Bigwood, Douglas
 ; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
 ; FILE REFERENCE: 5152
 ; CURRENT APPLICATION NUMBER: US/10/788,792
 ; PRIOR FILING DATE: 2004-02-27
 ; PRIOR APPLICATION NUMBER: US 60/450,655
 ; NUMBER OF SEQ ID NOS: 254
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 150
 ; LENGTH: 1464
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-788-792-150

Query Match 59.6%; Score 2765; DB 17; Length 1464;
 Best Local Similarity 55.4%; Pred. No. 4.8e-131;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;
 QY 1 GPPGEPGTGLPDPGGRGGSGRFGPCADGAGVAGKGPAGERSGPGAGKSGPGEA---57
 Db 179 GPMGSGPRGLPDPGAPGPGQFGQGPCEGPGASGPMGPRGPPGPKNGDDGEAGKP 238
 QY 58 GRCEAGLP---GAKGLTSGPSGPP-----DKTGTGPPAGODGRPPGPPGAR 105
 Db 239 GRPERGPPGQARGLPGTAGLPKMKHGRFGSLDGAAGDAGPAGKGPESGENGAP 298
 QY 106 QGAGVMPFGPKGAGSPGKAGER-----GVFPGPGAVGPA-----GKDGAGAG 150
 Db 299 GQMPRGLPGERGRPGAPGAPGARGNDGATCAAGPPGTGTGAPGPPGPPGAVGAKGAGPQ 358

QY 151 GPPGAPGAGERGEOGP-----AGSPGQGLPGPA 180
 Db 359 GPRGSGPQGVREGEPPGPAAGAPAGNPGADGQGAAGANGAPGAGPFPGARGPS 418
 QY 181 GPPGEAGKPEBQGVPGDLGAP-----GPSGPA-----GEPGPT 213
 Db 419 GPQGGPPGPKNGSCEPAGPSKGTGAKGPPGVGVQPPGPAAGECKRGARGEPGPT 478
 QY 214 GLPCEGREGGSGRFGPCADGAGVAGKGPAGERSGPGAGKSGPGEAGRGEAGLPAGAK 273
 Db 479 GLPFPGEKGGSGRFGPCADGAGVAGKGPAGERSGPGAGKSGPGEAGRGEAGLPAGAK 538
 QY 274 GLTSGPSGPPDGKTGPPGAPGQDGRPPGPPGARGQAGVWFGPKGAGEPKKAGER 333
 Db 539 GLTSGPSGPPDGKTGPPGAPGQDGRPPGPPGARGQAGVWFGPKGAGEPKKAGER 598
 QY 334 GVPFPAGVGPAGKDEAGAGQPPGAPGAGBRGEGGAGSPGFGGLPGPAGPPGAEAGKP 393
 Db 599 GVPFPAGVGPAGKDEAGAGQPPGAPGAGBRGEGGAGSPGFGFGQLFPGAPPPGAEAGKP 658
 QY 394 GEOGVFDLGAFCPSGAPGE-----PGP-----TGLPGPP 423
 Db 659 GEQGVFDLGAFCPSGARGGFGPGERGVQGGPAGPRGANGAPCNDGAKGDAGAPAP 718
 QY 424 GERGGSGRFGGADGAGVAGKGPAGERSGPGAGKSGPGE-----AGRGEAGLP 474
 Db 719 GSGGAPLQCMFGERGAAGLPKGDGADGPKGADGSPGKGVGLTGTPIGPPGAPAG 778
 QY 475 GAK-----GLTSGPSGPDGKTGPPAGODGRPPGPPGARGQ-----515
 Db 779 GDKGESGSPGAPGTGARGAPGDRGEPGPPGAGFAGPGADGQSGAKGEPGDAGAKDA 838
 QY 516 -----AGVWFGPKGAGEPKKAGERGVPPGPA-----VGPAGKDEAGAGQPP 561
 Db 839 GPPGAPGAPGPPPIGNVAPAKARGAGSGPAGTGFPGAAGRVGPPGPPSGNAGPPGP 898
 QY 562 GPAGP---AGERGEOGAGSPGFGGLPGPAGPPGAGKPGGQGVFCDLGAFCPSGAPAGEP 618
 Db 899 GPAGKEGGKPRGETGAPRGVEVGPAGGKSGADGAGAGTGTGPGIAGOR 958
 QY 619 GPTGLPGRGGRGSGRFGPCADGAGVAGKGPAGERSGPGAPG---KGSFGEAGRGEA 675
 Db 959 GVVGLPGORGERGFPGLFPGSGEPKQGPSCASGERGPPGPPGLAGPPGSGREGAP 1018
 QY 676 GLPGAKGLTSGPSGPDGKTGPPAGODGRPPGPPGARGQAGVWFGPKGAGEPKKAGER 735
 Db 1019 GAEGSGRDSGFAKEDRGETGAPGAPCAPGAPGVPVAGKSGDRGETGAPGAPGV 1078
 QY 736 KAGBRGVGPPGAVGAPKXGDE-----AGAQGPPG-----AGPA 771
 Db 1079 GPAGARGPAGPQGPREDKGETGEQDRIKGRHFGSLQGGPPGSGEQQSGASGPA 1138
 QY 772 GERGEOGAGSP---GFOGLPGPAGPGEAGKPGEQVFDLGAFCPSGAPAG 820
 Db 1139 GPRGPPGAGAPGKDLNGLPGLPPIGPPRGTGDAGVGPVPPGPPGPPG 1190

RESULT 11
 US-10-468-091-25
 ; Sequence 25, Application US/10468091
 ; Publication No. US20040157329A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ADP Pharmaceutical Pty Limited
 ; APPLICANT: The University of Sydney
 ; TITLE OF INVENTION: Matrix gene expression in chondrogenesis
 ; FILE REFERENCE: 500311
 ; CURRENT APPLICATION NUMBER: US/10/468,091
 ; CURRENT FILING DATE: 2003-08-13
 ; PRIOR APPLICATION NUMBER: AU P3116
 ; PRIOR FILING DATE: 2001-02-15
 ; NUMBER OF SEQ ID NOS: 42
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 25

Query Match 59.5%; Score 2761; DB 16; Length 1461;
Best Local Similarity 55.3%; Pred. No. 7.5e-131;
Matches 560; Conservative 31; Mismatches 229; Indels 192; Gaps 19;
; ORGANISM: Homo sapiens
US-10-468-091-25

QY 1 GPPGEPGTPGLPGRGSGSGFPCADQAVAGPKPAGRGSPGAGPKGSPGEA--- 57
DB 176 GPMGSGRGLPFPFAPGPGFQGPGEPEPEASGPMGRPPGPKNGDDGEAGKP 235
QY 58 GRPGEAGLP---GAKGLTSGSPGCP-----DGKTGPPGAGQDGRPPGPPGAR 105
DB 236 GRPGERGPPGQAGRLPGTAGLFGMKHGRGSLDGAKGDAGPAGPKGSPGSENGAP 295
QY 106 GQAGVMGPPGKGAAGEPKKAGER-----GVPPFGAVGPA-----GKDGEAGQ 150
DB 296 GQMGPRGLPGRGPFAGPAGAGNDGATGAAGPPTGTPAGPPGPFAGVAKGEAGPQ 355
QY 151 GPPGAPGAGRGQGP-----GPSGA-----GPGPT 213
DB 416 GPGGPGPPGPKNSGEPAGPSKDTGAKGEPGPGVQGPAGBEGKRGARGEFPT 475
QY 214 GLPDPGRRGPGSGRFGAGVAGPKPAGERSGPGAGPKGSPGAGRPGEAGLPKAX 273
DB 476 GLPDPGRRGPGSGRFGAGVAGPKPAGERSGPGAGPKGSPGAGRPGEAGLPKAX 535
QY 274 GLTSGSPGPDGKTGTPGPGAGQDGRPPGPPGARGQAGVMGPPKGAAGEPKKAGER 333
DB 536 GLTSGSPGPDGKTGTPGPGAGQDGRPPGPPGARGQAGVMGPPKGAAGEPKKAGER 595
QY 334 GVPDPGAVGPKDGEAGAGPPGAGPAGERGEQAGSPGQGLPGRAGPGEAGKP 393
DB 596 GVPDPGAVGPKDGEAGAGPPGAGPAGERGEQAGSPGQGLPGRAGPGEAGKP 655
QY 394 GQGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
DB 656 GQGVPGDLGAPGSPGAGERGFPGERGVQGPAGPRGANGAGNDGAKGDAGAPG 715
QY 424 GERGPGSRGPPGAGVAGPKPAGERSGPGAGPKGSPGE-----AGRPGEAGLP 474
DB 716 GSGGAPGLQGMGERGAAAGLPKPKDGRDAGPKGADGSGFKDGVRLTGTPTGPPGAGP 775
QY 475 GAK-----GLTSGSPGPDGKTGTPGPGAGQDGRPPGPPGARGQ----- 515
DB 776 GDKGSGSPGAPGTGARGAPDGRGEPGPPGAGPAGPPGAGPPGAGKGEFGDAGAKDA 835
QY 516 -----AGVMGFPKPKGAAGEPKKAGERGVPPGPA-----VGPAGKDEAGAGQPP 561
DB 836 GPPGAPGAPGPPGPIGNVAPKAGKARGSPGATGFFGAAGRVPSPGSGNAGPPGP 895
QY 562 GPAGP-----AGERGEQAGSPGQGLPAGPAGPAGKPGQGVPGDLGAPGSPGAGEP 618
DB 896 GPAGKEGKGPRGETGPAGRPGEVPPGPPGAGEKSGFGADGPAGAPCTGPGQIAGQR 955
QY 619 GPTGLPGRGPGSGSRGFPFAGDGVAGPKPAGERSGPPGAGP-----KXSGEAGRPGEA 675
DB 956 GVVGLPQQRGERGFPCLPSPSGPKQKPSGASGERGPPGPMGPPGLAGPPGSGREGAP 1015
QY 676 GLPAGKLTGSPGSPGPDGKTGTPGPGAGQDGRPPGPPGARGQAGVMGFPKPKGAAGEP 735
DB 1016 GAEGSPGRDGSFCAKGRGETGAPGPPGAPGAPGFPVGPAGKSGDRGETGAPGAPGV 1075
QY 736 GKAGRGVPGPVGAVGPAKDE-----AGAGPPGP-----AGPA 771
DB 1076 GVPAGKAPGAPGPRGDKGETGEQDGRGIKHGRGSGLGPPGPPGSPGEGQSPGASGPA 1135

QY 772 GERGEQGPAGSP---GFQGLPAPAGPPGGEAGKPKGQGVPGDLGARGPSGPAG 820
DB 1136 GPRGPPGAGARKDGLNGLPGIPGPRGRGTGDAGPVGPPGPPGPPG 1187

RESULT 12

US-10-291-265-243
; Sequence 243, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al.
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-243

Query Match 59.5%; Score 2761; DB 14; Length 1464;
Best Local Similarity 55.3%; Pred. No. 7.5e-131;
Matches 560; Conservative 31; Mismatches 229; Indels 192; Gaps 19;

QY 1 GPPGEPGTPGLPGRGSGSGFPCADQAVAGPKPAGRGSPGAGPKGSPGEA--- 57
DB 179 GPMGSGRGLPFPFAPGPGFQGPGEPEASGPMGRPPGPKNGDDGEAGKP 238
QY 58 GRPGEAGLP---GAKGLTSGSPGCP-----DGKTGPPGAGQDGRPPGPPGAR 105
DB 239 GRPGERGPPGQAGRLPGTAGLFGMKHGRGSLDGAKGDAGPAGPKGSPGSENGAP 298
QY 106 GQAGVMGFPKPKGAAGEPKKAGER-----GVPPFGAVGPA-----GKDGEAGQ 150
DB 299 GQMGPRGLPGRGPFAGPAGAGNDGATGAAGPPTGTPAGPPGPFAGVAKGEAGPQ 358
QY 151 GPPGAPGAGRGQGP-----GPSGA-----AGSPGQGLP 180
DB 359 GPRGSEGPQVGRGEPFPGPAGAAAGPAGNPGADQGPAGKANGAPGIAGPFGARGPS 418
QY 181 GPPGEPGTPGLPGRGSGSGFPCADQAVAGPKPAGERSGPPGAGPKGSPGEA--- 213
DB 419 GPGGPGPPGPKNSGEPGAPGSKGTGAKGEPGVVQGPAGBEGKRGARGEFPT 478
QY 214 GLPDPGRRGPGSGRFGAGVAGPKPAGERSGPPGAGPKGSPGAGRPGEAGLPKAX 273
DB 479 GLPDPGRRGPGSGRFGAGVAGPKPAGERSGPPGAGPKGSPGAGRPGEAGLPKAX 538
QY 274 GLTSGSPGPDGKTGTPGPGAGQDGRPPGPPGARGQAGVMGFPKPKGAAGEPKKAGER 333
DB 539 GLTSGSPGPDGKTGTPGPGAGQDGRPPGPPGARGQAGVMGFPKPKGAAGEPKKAGER 598
QY 334 GVPDPGAVGPKDGEAGAGPPGAGPAGERGEQAGSPGQGLPGRAGPGEAGKP 393
DB 599 GVPDPGAVGPKDGEAGAGPPGAGPAGERGEQAGSPGQGLPGRAGPGEAGKP 658
QY 394 GQGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
DB 659 GQGVPGDLGAPGSPGAGERGFPGERGVQGPAGPRGANGAGNDGAKGDAGAPGAP 718
QY 424 GERGPGSRGFPFAGDGVAGPKPAGERSGPPGARGPSGPE-----AGRPGEAGLP 474

```

Db 719 GSGAGLQGMGERGAAGLPGKGDAGPKGADGSGKDGKGLTGP:GPDPAGAP 778
QY 475 GAK-----GLTSGPSGPDGKTGPPGAGDGRPPGPPGARGQ-----515
Db 779 GDKGESGSPGAPGTCARGAPGDRGEPGPPGAPGAPGADGQFGAKGPDAGAKGDA 838
QY 516 -----AGVMGFRPKGAAGEBKACGERGVPPGGA-----VGPAGKDEAGAGPP 561
Db 839 GPFGAGPAGPPEPTGNVGAFAKAGSAGPFGATGPGAGRVGPPGPGSGNAGPPCP 898
QY 562 GPAGP-----AGRGEGGAGSPGQFLGPPAGPPGAGKPGGQVPGDILGAFGSPGAGEP 618
Db 899 GPAGKEGKGPRGETGAPRGVGVPPGPPGAGEKSGFADGAPAGTGTGPGIAGQR 958
QY 619 GPTGLPVPGERGGSGRFGCADGVAGPKGAPGERSGPPAGP---KSGPGEAGRGEA 675
Db 959 GVUGLPGORGEKGFELPSPGEPKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAP 1018
QY 676 GLPGAAGLTGSPGSPGDKTGPAGQDGRPPGPPGARGQAGVMGFPKGAAGEP 735
Db 1019 GAEGSPGRDGSFAGKSDRGETGAPGPPGAPGAPGAPGVGVPAGKSDRGETGAPGAPV 1078
QY 736 GKAGERGVPPGAVGPAKGDGE-----AGAQGPPG-----AGPA 771
Db 1079 GPVGARGPAGPQGRGDXGETGEQDGRGKGRHGFSLGQSPGPPGSPGSGASGPA 1138
QY 772 GERGQGPAGSP---GFQGLPQAPGPEAGKPGEGQVPGDLGAPGSPGAP 820
Db 1139 GPRGPGSAGAPKDLNGLPGTGPGRGRTGDAGVGPVGGPPGPPG 1190

```

RESULT 13

```

US-10-104-889-16
; Sequence 16, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVERTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8494
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

```

;
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-10-104-889-16

```

```

Query Match 59.5%; Score 2759; DB 15; Length 1057;
Best Local Similarity 55.3%; Pred. No. 7,5e-131;
Matches 560; Conservative 31; Mismatches 229; Indels 192; Gaps 19;

```

```

QY 1 GPPEGPTGLPFGPGERGCGSRGFPFGADGVAGPKGAGERGSPGAPGKSGPGEA---57
Db 18 GPMGFSGRGLPGPPGAPGPPGQFPQFPBPBPFGCAGPMGPRGPPGPPKNGDDCEAGKP 77
QY 58 GRPEAGLP---CAKLTGSPGSPGP-----DGKTGPPGAPGQDGRPPGPPGAP 105
Db 78 GRPGERGPPGQCARGLPOTAGLPGKVRHGRGSLUDGAKGADGAPGKSGPGEAGAP 137
QY 106 GQAGVMGFPFKGAAGEFGKAGER-----GVFPFGGAVGPA-----GKDEAGAP 150
Db 138 GQMGPRGLPGRGRPGAPGAPGARGNDGATGAAGPPGTPGAPGPPGCAVCAKGEAGP 197
QY 151 GPFGPAGPAGERGEGP-----AGSPGFQQLPAPA 180
Db 198 GPRGSEGPQGVGRGFPFPAGAGPAGNPGADGQPGAKGANGAPGAGPFGFARGPS 257
QY 181 GPPGEAGKPGEGQVPGDLGAP-----GPSGPA-----GERGPT 213
Db 258 GPQGGPFPKNSGSEFGAPGSKDGTAKGEPGVGVGPPGAGEGKRGARGPPT 317
QY 214 GLPFGPGERGSGRFPFGCADGVAGPKGAPGERSGPPGAPGKSGFGEAGRPGEAGLP 273
Db 318 GLFPGPGERGSGRFPFGCADGVAGPKGAPGERSGPPGAPGKSGFGEAGRPGEAGLP 377
QY 274 GLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQAGVMGFPKGAAGEGPKAGER 333
Db 378 GLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQAGVMGFPKGAAGEGPKAGER 437
QY 334 GVPFGPAGVGPAGKDEAGAGQPPGAPGAGERGQSGPAGSPGQGLFGPAGPFGGAGKP 393
Db 438 GVPFGPAGVGPAGKDEAGAGQPPGAPGAGERGQSGPAGSPGQGLFGPAGPFGGAGKP 497
QY 394 GEGVPGDLGAPGSPGAGE-----PGP-----TGLFGPP 423
Db 498 GEQGVPGDLGAPGSPGAGERGFPGERGVQPPGAPGPRGANGAPGNDGAKGDAGAP 557
QY 424 GERGGPGRGPPGADGVAGPKGAPGERSGPPGAPGKSGPGE-----AGRGEGAGLP 474
Db 558 GSQGAPGLQGMFGERGAAGLPGKDRGADGPKGADGSPGKDGVRGLTGTPTGPPGAP 617
QY 475 GAK-----GLTSGSPGSPGDKTGPAGQDGRPPGPPGARGQ-----515
Db 618 GDKGESGSPGAPGTCARGAPGDRGEPGPPGAPGAPGADGQFGAKGPDAGAKGDA 677
QY 516 -----AGVMGFRPKGAAGEBKACGERGVPPGGA-----VGPAGKDEAGAGPP 561
Db 678 GPFGAGPAGPPEPTGNVGAFAKAGSAGPFGATGPGAGRVGPPGPGSGNAGPPCP 737
QY 562 GPAGP-----AGRGEGGAGSPGQFLGPPAGPPGAGKPGGQVPGDILGAFGSPGAGEP 618
Db 738 GPAGKEGKGPRGETGAPRGVGVPPGPPGAGEKSGFADGAPAGTGTGPGIAGQR 797
QY 619 GPTGLPVPGERGSGRFGCADGVAGPKGAPGERSGPPAGP---KSGPGEAGRGEA 675
Db 798 GVUGLPGORGERGPPGLPFGSGEFKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAP 857
QY 676 GLPGAAGLTGSPGSPGDKTGPAGQDGRPPGPPGARGQAGVMGFPKGAAGEP 735
Db 858 AAEPSGRDGSFAGKSDRGETGAPGPPGAPGAPGAPGVGVPAGKSDRGETGAPGAPV 917
QY 736 GKAGERGVPPGAVGPAKGDGE-----AGAQGPPG-----AGPA 771
Db 918 GPAGARGPAGPQGRGDXGETGEQDGRGKGRHGFSLGQSPGPPGSPGSGASGPA 977

```


Db 359 GPRGSEGGVGEPPGPPGAGAGAGNPNAGDQPGAKGANGARFIAGAFGFFGARGPS 418
QY 181 GPPGEAGKPEQGVGDLGAP-----GPSGPA-----GEPGPT 213
Db 419 GQGGGPPGPKGNSGEPGAPGSGKDTGAKGEPGVGVGGPPGAGEGKRGARGEPFT 478
QY 214 GLPGPPGERGGSGGFFGADGAVAGPKGPAGRGSPGAGKSGPGEAGRPGEAGLPKAK 273
Db 479 GLPGPPGERGGSGGFFGADGAVAGPKGPAGRGSPGAGKSGPGEAGRPGEAGLPKAK 538
QY 274 GLTSGSGSPGPKGTGPPGAGQDGRPGPPGARGOAGVMGPPGKGAAGEPKKAGER 333
Db 539 GLTSGSGSPGPKGTGPPGAGQDGRPGPPGARGOAGVMGPPGKGAAGEPKKAGER 598
QY 334 GVPGGPAGVGPAGKDEAGAGQPPGPPGAGRGSPGAGSPGFGQGLFPAGPPGEAGKP 393
Db 599 GVPGGPAGVGPAGKDEAGAGQPPGPPGAGRGSPGAGSPGFGQGLFPAGPPGEAGKP 658
QY 394 GEQGVPGDLGAPSGPAGE-----PGP-----TGLPGPP 423
Db 659 GEQGVPGDLGAPSGPAGE-----PGP-----TGLPGPP 718
QY 424 GERGGSGRGGFADGVAGPKGPAGRGSPGAGKSGPGE-----ACRPGAGLP 474
Db 719 GSQGAFLGQMGGERGAAGLPKPKDGRDAGPKGADGSPGKDGVRGLTGPFGPPGAP 778
QY 475 GAK-----GLTSGSGSPGPKGTGPPGAGQDGRPGPPGARGQ-----515
Db 779 GSKGSGSGPAGTGPAGRGDRGPPGPPGAGPPGADGQFGAKGEPGDAGAKGDA 838
QY 516 -----AGVMGPPGKGAAGEPKKAGERGVPPGGA-----VGPAGKDEAGAGQPP 561
Db 839 GPPGAGPAGPPGPIGNVGAAGKAGARSAGPPGATGPPGAAGRVGPPGSGNAGPPGPP 898
QY 562 GPAGP---AGERGQGPAGSPGQGLPQAPGPPGEPAGKPGQGVGDLGAPGSPGAGEP 618
Db 899 GPAGKEGKGPRGETGPRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 958
QY 619 GFTGLPGLPGERGGSGRFFGADGAVAGPKGPAGRGSPGAPG---KSGPGEAGRPGEA 675
Db 959 GVVLPGQRGERGFFGLPSPGEPKQGPAGSGRPPGPPGPPGPPGPPGPPGPPGPPGPP 1018
QY 676 GLPGAKGLTSGSPGPKGTGPPGAGQDGRPGPPGARGOAGVMGPPGKGAAGEP 735
Db 1019 AAGSGSPGRDGSAGKDRGETGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 1078
QY 736 GKAGERGVPPGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 771
Db 1079 GPPGARGPAGPQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1138
QY 772 GERGGGPPAGSP---GFGGLPAGPAGPPGAGKPGQGVGDLGAPGPPGAPG 820
Db 1139 GPRGPPGAGAPKDLNGLPGLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1190

Search completed: October 18, 2004, 13:58:35
Job time : 112.622 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:31:10 ; Search time 32.5431 Seconds
(without alignments)
2427.363 Million cell updates/sec

Title: US-10-658-989A-3
Perfect score: 4640
Sequence: 1 GPGGPGTGLGPPGCGG.....GEQGVPGDLGAPGSPGAG 821
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: Pirl:.*
2: Pirl:.*
3: Pirl:.*
4: Pirl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2755	59.4	1464	1 CGHU1S	collagen alpha 1(I)
2	2732.5	58.9	1453	2 S21626	collagen alpha 1(I)
3	2679	57.7	1042	1 CGHU1S	collagen alpha 1(I)
4	2477	53.4	1418	2 T45467	collagen alpha 1(I)
5	2476	53.4	1487	1 CGHU6C	collagen alpha 1(I)
6	2460	53.0	1419	2 A41182	collagen alpha 1(I)
7	2460	53.0	1487	2 B41182	collagen alpha 1(I)
8	2401	51.7	1486	1 B40333	collagen alpha 1(I)
9	2389	51.5	1492	2 A40333	collagen alpha 1(I)
10	2379	51.3	1049	1 CG807S	collagen alpha 1(I)
11	2377.5	51.2	1464	2 S59856	collagen alpha 1(I)
12	2370	51.1	1466	1 CGHU7L	collagen alpha 1(I)
13	2326	50.1	1496	1 CGHU2V	collagen alpha 2(V)
14	2303.5	49.6	1497	2 I49607	procollagen type V
15	2300	49.6	1373	1 A43291	collagen alpha 2(I)
16	2263.5	48.8	1366	1 CGHU2S	collagen alpha 2(I)
17	2235	48.2	671	1 CGHT1S	collagen alpha 1(I)
18	2229.5	48.0	1838	1 CGHU1V	collagen alpha 1(V)
19	2208.5	47.6	1843	2 S18803	collagen alpha 1(V)
20	2193.5	47.3	779	1 CG801S	collagen alpha 1(I)
21	2169	46.7	1806	1 CGHU1E	collagen alpha 1(X)
22	2161.5	46.6	886	2 I50694	collagen alpha 1(I)
23	2136.5	46.0	1414	2 S23809	collagen alpha 2(I)
24	2074.5	44.7	1027	2 S28774	collagen alpha 1(V)
25	2071	44.6	2944	2 A54849	collagen alpha 1(V)
26	2060.5	44.4	1546	1 CGHU2E	collagen alpha 2(X)
27	2058	44.1	1691	1 CGHU1B	collagen alpha 5(I)
28	2046	44.1	1690	1 CGHU1B	collagen alpha 4(I)
29	2036	43.9	1549	2 I48103	type VII collagen

RESULT 1
CGHU1S

collagen alpha 1(I) chain precursor - human

N;Alternate names: procollagen alpha 1(I) chain

C;Species: Homo sapiens (man)

C;Date: 12-Aug-1981 #sequence revision 04-Oct-1996 #text change 09-Jul-2004

C;Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S1

5269; A29439; I53466; A02852; I37247

R;D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J

Gene 67, 105-115, 1988

A;Title: Complete nucleotide sequence of the region encompassing the first twenty-five

A;Reference number: I60114; MUID:88329734; PMID:2843432

A;Accession: I60114

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-369, 'L', 371-589 <DAL>

A;Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNI

Biochem. J. 253, 919-922, 1988

A;Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human

A;Reference number: S01143; MUID:89025644; PMID:3178743

A;Accession: S01143

A;Molecule type: mRNA

A;Residues: 1-472 <TRO>

A;Cross-references: EMBL:X07884; NID:G30015; PIDN:CAA30731.1; PID:G30016; GB:M36546; NI

A;Note: Submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C

Nature 310, 337-340, 1984

A;Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation o

A;Reference number: A93335; MUID:84270697; PMID:6462220

A;Accession: A93335

A;Molecule type: DNA

A;Residues: 1-58, 'Q', 60-181 <CHU>

A;Cross-references: EMBL:X00820; NID:G35657; PIDN:CAA2394.1; PID:G35658

R;Roscow, C.M.S.; Vergeer, W.P.; du Picoy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W,

J. Biol. Chem. 262, 15151-15157, 1987

A;Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene en

A;Reference number: I55254; MUID:88033098; PMID:2822714

A;Accession: I55254

A;Status: translation not shown; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-45 <ROS>

A;Cross-references: GB:J02829; NID:G180387; PIDN:AAAS1993.1; PID:G180388

R;Bornstein, P.; McKay, J.; Morishima, J.K.; Devareyalu, S.; Gelinas, R.E.

Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987

A;Title: Regulatory elements in the first intron contribute to transcriptional control

A;Reference number: A39943; MUID:88097389; PMID:3480516

A;Accession: A39943

A;Molecule type: DNA

A;Residues: 1-34 <BOR>

A;Cross-references: GB:J03559; NID:G180876; PIDN:AAAS2052.1; PID:G553238

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
 A;Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
 A;Reference number: 155237; MUID:85130970; PMID:2857713
 A;Accession: 155237
 A;Status: translation not shown; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-34 <CH2>
 A;Cross-references: GB:M10627; NID:G180383; PIDN:AAA5192.1; PID:9553226
 R;Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollis
 J. Biol. Chem. 265, 6312-6317, 1990
 A;Title: In vivo and in vitro noncovalent association of excised alpha 1(I) amino-termina
 some type VII.
 A;Reference number: A35233; MUID:90202908; PMID:2318855
 A;Accession: A35233
 A;Molecule type: protein
 A;Residues: 33-52 <WR>
 A;Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
 R;Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
 EMBO J. 8, 1705-1710, 1989
 A;Title: A base substitution in the exon of a collagen gene causes alternative splicing
 A;Reference number: S09400; MUID:89356643; PMID:2767050
 A;Accession: S09400
 A;Molecule type: mRNA
 A;Residues: 156-183 <WEI>
 R;Click, E.M.; Bornstein, P.
 Biochemistry 9, 4699-4706, 1970
 A;Title: Isolation and characterization of the cyanogen bromide peptides from the alpha
 A;Reference number: A30567; MUID:71038625; PMID:5529814
 A;Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
 A;Accession: B90567
 A;Molecule type: protein
 A;Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'
 A;Experimental source: skin
 A;Note: evidence for 170-allysine
 R;Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, B
 Eur. J. Biochem. 192, 153-159, 1990
 A;Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
 A;Reference number: S11372; MUID:90382436; PMID:2169412
 A;Accession: S11372
 A;Molecule type: protein
 A;Residues: 173-187; 274-287, 'P', 289 <BAE>
 A;Note: sequence of collagen alpha 1(S) (I) isolated from bone after pepsin digestion
 R;Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
 J. Biol. Chem. 266, 21827-21832, 1991
 A;Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
 cooperative melting of intact type I collagen.
 A;Reference number: 155342; MUID:92042092; PMID:1718984
 A;Accession: 155342
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 258-268; 1347-1357 <DEA>
 A;Cross-references: GB:S67495; NID:G239007; PIDN:AAB20350.1; PID:G239008
 A;Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
 R;Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
 J. Biol. Chem. 245, 5042-5048, 1970
 A;Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
 A;Reference number: A92069; MUID:71001508; PMID:4319110
 A;Accession: A92069
 A;Molecule type: protein
 A;Residues: 263-268 <MOR>
 A;Experimental source: skin
 A;Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
 R;Labhard, M.E.; Hollister, D.W.
 Matrix 10, 124-130, 1990
 A;Title: Segmental amplification of the entire helical and telopeptide regions of the cd
 A;Reference number: S15989; MUID:90326017; PMID:2374517
 A;Accession: S15989
 A;Molecule type: mRNA
 A;Residues: 281-302; 402-420; 823-843; 925-944; 1026-1045; 1143-1162 <LAB>
 R;Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N
 Connect. Tissue Res. 29, 1-11, 1993
 A;Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
 A;Reference number: 152905; MUID:93339042; PMID:8339541

A;Accession: 152905
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 342-352, 'C', 354-359 <WI2>
 A;Cross-references: GB:S64717; NID:G408195; PIDN:AAB27677.1; PID:G408196
 A;Note: mutant sequence from patient with osteogenesis imperfecta
 R;Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Ekemberry, E.F.; Prockop, D.J.
 Biochemistry 22, 5213-5223, 1983
 A;Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha
 A;Reference number: A90476; MUID:84080385; PMID:6689127
 A;Accession: A90476
 A;Molecule type: mRNA
 A;Residues: 425-1250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>
 A;Cross-references: GB:K01228; NID:G180391; PIDN:AAA5195.1; PID:G180392
 A;Note: sequence partially completed for missing nucleotides by A39439
 R;Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
 J. Biol. Chem. 260, 691-694, 1985
 A;Title: Multixon deletion in an osteogenesis imperfecta variant with increased type I
 A;Reference number: A22161; MUID:85104934; PMID:2381843
 A;Accession: A22161
 A;Molecule type: DNA
 A;Residues: 472-594, 'R', 596-607 <CH3>
 A;Cross-references: GB:K03178; GB:K03179; NID:G179612; NID:G179613; PIDN:AAA51847.1; PID
 A;Note: the authors translated the codon CGT for residue 595 as Pro
 R;Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
 Am. J. Hum. Genet. 46, 1034-1040, 1990
 A;Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained
 A;Reference number: A35336; MUID:90252792; PMID:2339700
 A;Accession: A35336
 A;Molecule type: mRNA
 A;Residues: 710-720, 'E', 722-737, 'E', 739-745 <WAL>
 A;Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
 R;Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes
 Hum. Mol. Genet. 3, 2201-2206, 1994
 A;Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the c
 A;Reference number: 154365; MUID:95187161; PMID:7881420
 A;Accession: 154365
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 746-766, 'S', 768-781 <FOR>
 A;Cross-references: GB:L47667; NID:G1009093; PIDN:AAB59576.1; PID:G1009094
 R;Chesler, S.D.; Wallis, G.A.; Byers, P.H.
 J. Biol. Chem. 268, 18218-18225, 1993
 A;Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
 A;Reference number: A47426; MUID:93352646; PMID:8349697
 A;Accession: A47426
 A;Molecule type: mRNA
 A;Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>
 A;Cross-references: GB:S84596; NID:G407589; PIDN:AAB27856.1; PID:G407590
 A;Note: sequence extracted from NCBI backbone (NCBIN:136444, NCSTIP:136445)
 A;Note: does not represent an experimentally determined sequence but three different mut
 A;Accession: B47426
 A;Molecule type: mRNA
 A;Residues: 1179-1464 <CH4>
 A;Experimental source: normal dermal fibroblast culture
 A;Accession: C47426
 A;Molecule type: mRNA
 A;Residues: 1179-1276, 'H', 1278-1464 <CH5>
 A;Experimental source: fetal cell 86-237
 A;Accession: D47426
 A;Molecule type: mRNA
 A;Residues: 1179-1336, 1339-1464 <CH6>
 A;Experimental source: fetal cell 86-146
 A;Accession: E47426
 A;Molecule type: mRNA
 A;Residues: 1179-1387, 'R', 1389-1464 <CH7>
 R;Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Nid
 J. Biol. Chem. 263, 14605-14607, 1988
 A;Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptide c
 A;Reference number: 155269; MUID:89008319; PMID:3170557
 A;Accession: 155269

A;Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1187-1194, 'C', 1196-1200 <COH>
A:Cross-references: GB:M23213; NID:G494042; PIDN:AAB59363.1; PID:G499622
A:Note: mutant sequence from a patient with mild osteogenesis imperfecta
R;Mackelae, J.K.; Raessina, M.; Virta, A.; Vuorio, E.
Nucleic Acids Res. 16, 349, 1988
A:Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

Query Match 59.4%; Score 2756; DB 1; Length 1464;
Best Local Similarity 55.3%; Pred. No. 3.5e-130;
Matches 560; Conservative 31; Mismatches 229; Indels 192; Gaps 19;

QY 1 GPPGEPPTGLPGLGPPGGRGSGRFPFGADGAVGKPGAGERSGPGAGPKGSPFGEA--- 57
DB 179 GPMGSGRGLPGLPAGPGPGQFGPGEPCEPCASGPMGPRGPGPKGNGDGEAGKP 238
QY 58 GRPEAGLPL---GAKGLTSGSGSP-----DGKTGPPGAGODGPPGPPGAP 105
DB 239 GRPGERGPPGQAGRLPGLTGLGPKMGRHGFSLDGAKDAGPAGPKGSPGPGENGAP 298
QY 106 GQAGVMGFPFGKGAAGEPKAGER-----GVPPFGAVGPA-----GKDGGAQA 150
DB 239 GQMPRGLPGRGPRGAPGAPAGARGNDGATGAAGPPGTGPPGPPGAVGAKGEAGQP 358
QY 151 GPPGAPGAGRGSGP-----AGSPGQGLPCPA 180
DB 359 GPRGSEGPQVRGPPGPPGAGAGPAGNPGADGQCGAKGANGAPGAPGPPGARGPS 418
QY 181 GPPGAGKPGQGVPGDLGAP-----GPSGA-----GEPGPT 213
DB 419 GQGGGPPGPKNGSGEPGAKGSDTGAKGEPGVQGGPPGAGBEKRGARGEPT 478
QY 214 GLPGLPGRGPGSGRFPFGADGAVGKPGAGERSGPGAGPKGSPFGAGRPGAGLPGAK 273
DB 479 GLPGLPGRGPGSGRFPFGADGAVGKPGAGERSGPGAGPKGSPFGAGRPGAGLPGAK 538
QY 274 GLTSGPSPGPDGKTGPPGAGODGRPPGARGQAGVMGPPGKGAAGEPKGAGER 333
DB 539 GLTSGPSPGPDGKTGPPGAGODGRPPGARGQAGVMGPPGKGAAGEPKGAGER 598
QY 334 GVPGLPAGVAPGAKDGEAGAGQPPGAPGAGERSGPGAGFPGLPFPAGPGEAGKP 393
DB 599 GVPGLPAGVAPGAKDGEAGAGQPPGAPGAGERSGPGAGFPGLPFPAGPGEAGKP 658
QY 394 GQGVPGDLGAPGSGPAGE-----PGP-----TGLPSP 423
DB 659 GQGVPGDLGAPGSGPARGERFPGERGVQGGPPGAPRGANGAPGNDGAKGDAGAPAP 718
QY 424 GERGGPSRFPFGADGAVGKPGAGERSGPGAGPKGSPGE-----AGRPGEAGLP 474
DB 719 GSQAPGLQMPGREGAAGLPKPKDRGDAGPKGADGSGKDGVRGLTGPGLPFPAGAP 778
QY 475 GAK-----GLTSGPSGPDGKTGPPGAGODGRPPGPPGARGQ----- 515
DB 779 GDKGESGSGPAGPTGARGAPGDRGEPGPPGAPGAGPAGDQGPAGKGPAGDAGAKGDA 838
QY 516 -----AGVMGFPGPKAAERKAGERGVPGPB-----VGPAGKGEAGAGQPP 561
DB 839 GPPGAPGAGPPGPIGVNAPGAKGARGAGSPGATGPPGAAGRVGPPGSGNAGPPPP 898
QY 562 GPAGP---AGERGEQGPAGSPFGILPFGAGPFGAGKFGEGQVFGDLGAPGSPGAP 618
DB 899 GPAGKGGKPRGETGPRGPEVGVPPGPPGAGEKSPGADGAPAGPTPGPQGIAGR 958
QY 619 GPTGLPGRGPGGSRFPFGADGAVGKPGAGERSGPPGAP---KSPGCEAGRPGEA 675
DB 959 GVVGLPQGRGERGFPGLPSPGSPGKQPSGASGERGPPGPPGPPGLAGPPGESREGAP 1018
QY 676 GLPAGKGLTSGPSGPDGKTGPPGAGODGRPPGPPGAGQAGVMGFPFGKGAAGEP 735
DB 1019 GAEGPSRDPGSPGAKGDRGETGAPGPPGAXGAGPVGPRGKSGDRGETGAPGAPGV 1078

QY 736 KGAGRGVPPGPGAVGAPKADGE-----AGQGGPPG-----AGPA 771
DB 1079 CPAGARGPAGPQPRGDKGETGEGDRIKGRHGFSGGLQGGPPGPGSGCEQGPSASGPA 1138
QY 772 GRRGSGGAGS---GFGGLPAGPAGPGEAGKPGGQGVFGDLGAPGSPGAP 820
DB 1139 GRRGPGGAGAPCKGLNGLPGLPFPGRGTGAGVPGPPGPPGPPG 1190

RESULT 2
S21626
collagen alpha 1(I) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text change 09-Jul-2004
C;Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626
R;Li, S.W.; Khillan, J.; Prockop, D.J.
Matrix Biol. 14, 593-595, 1994
A;Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I
A;Reference number: S57243
A;Accession: S57243
A:Molecule type: mRNA
A:Residues: 1-1453 <LIS>
A:Cross-references: UNIPROT:P11087; EMBL:U08020; NID:G470673; PIDN:AAA88912.1; PID:G470
R;Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs
A;Reference number: S16176; MUID:91274355; PMID:2054384
A;Accession: S16374
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1453 <MET>
A:Cross-references: EMBL:X57981; NID:G50484; PIDN:CAA41046.1; PID:G50485
R;French, B.T.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A;Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A;Reference number: A23982; MUID:86137403; PMID:3841523
A;Accession: A23982
A:Molecule type: mRNA
A:Residues: 518-1128 <PRE>
A:Cross-references: GB:M14423; NID:G192261; PIDN:AAA37333.1; PID:G192262
R;Monson, J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A;Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for
A;Reference number: I49559; MUID:83141374; PMID:6298597
A;Accession: I49559
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 735-1130 <RES>
A:Cross-references: GB:M17491; NID:G192263; PIDN:AAA37334.1; PID:G192264
R;Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A;Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads
A;Reference number: I49557; MUID:84170331; PMID:6324198
A;Accession: I49557
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-25 <RES>
A:Cross-references: GB:X01688; NID:G192246; PIDN:AAA37330.1; PID:G553881
R;Fenton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A;Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A;Reference number: S39789; MUID:94092741; PMID:8268229
A;Accession: S39789
A:Molecule type: DNA
A:Residues: 1-80, 'E', 82-105, 'D', 107-185, 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-
R;Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
Mol. Cell. Biol. 14, 5950-5960, 1994
A;Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indire
A;Reference number: I48300; MUID:94344105; PMID:8065328
A;Accession: I48300
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>

A;Cross-references: EMBL:X54876; NID:G50486; PIDN:CAA38657.1; PID:G50487

C;Genetics:

A;Gene: COL1A1
A;Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/3
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-151/Domain: amino-terminal propeptide #status predicted <PRO>
F;30-89/Domain: von Willebrand factor type C repeat homology <VWC>
F;152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>
F;1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <CC>

Query Match 58.9%; Score 2732.5; DB 2; Length 1453;

Best Local Similarity 54.7%; Pred. No. 5.1e-129;

Matches 547; Conservative 33; Mismatches 240; Indels 180; Gaps 16;

QY 1 GPPGPGPTGLPGPGE-----RGPGSRGPPGADGAVAGKPGAGERSGPPA 48

DB 180 GPPGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 239

QY 49 GPKSGEAGRP-----GEAGLPGAKGLTSGSPGPGPGKTPGPPGAGQD 93

DB 240 GARGLPGTAKLPMKXKRGFSGLDAKGDAGPAGKPGGPGGPGGPGGPGGPGG 299

QY 94 GRPGPPGPPGARGQAGVMGFPGPKGAAGE-----PGKAGRGVPPGPGVAGPKDGEA 147

DB 300 GRPGPPGTAGRGNDGAVGAAGPPGPTGTPGPPGAVGAKGAGPGQAGSGSPQGV 359

QY 148 GAGGPPGPPGAGRGEGQGP-----AGSPGQGLPGPAGPPGEGAGKPGCEQ 192

DB 360 GPGGPPGAGAAAGPAGNPGADQPGKAGANGAPGIAGAPGPPGAGSPGSPGPGPK 419

QY 193 GVPGLDGLAP-----GPSGA-----GPPGFTGLPGPPGRRGGP 225

DB 420 GNSGPGAPGNKGDGAKGECATGVQPPGPPAGEEGRKRGAGFPGSGLEPPGRRGGP 479

QY 226 GSRGPPGADGAVGPKGPPGAGRGSPGAPKSPGEGAGRPBAGLPGAKGLTSGSPGCPD 285

DB 480 GSRGPPGADGAVGPKGPPGAGRGSPGAPKSPGEGAGRPBAGLPGAKGLTSGSPGCPD 539

QY 286 GKTGPPGAGQDGRGPPGPPGARGQAGVMGFPGPKGAAGSPGKAGRGVPPGPGVGA 345

DB 540 GKTGPPGAGQDGRGPPGPPGARGQAGVMGFPGPKGAAGSPGKAGRGVPPGPGVGA 599

QY 346 GKDGAAGQPPGPPGAGRGEGGPPGAGSPGQGLPGPAGPPGAGKPGGQVPGDILGAP 405

DB 600 GKDGAAGQPPGPPGAGRGEGGPPGAGSPGQGLPGPAGPPGAGKPGGQVPGDILGAP 659

QY 406 GPSGPGAG-----RGP-----TGLPGPPGRRGGPGSRGFP 435

DB 660 GPSGAGRGPPGRRGVGPPGPPGAGRGNGNAPNDGAKGDTGAPGAPGSGQAGFLQGMP 719

QY 436 GADGAGKPGAGRGSPGPPGKSGPGE-----AGRPBAGLPGAKGLTSGSPG 486

DB 720 GERGAAGLPGPKGRDAGPKGADGSPKDGARGLGTGPIGPPGAGAGKPGGAGSGPP 779

QY 487 GPD-----GKTGPPGAGQDGRGPPGPPGARGQ-----AGVMGFP 522

DB 780 GPTGARGAPDRGAGPPGAGFAGFPAGDQPGKAGGPGDTGVKGDAGPPGAGPAGPP 839

QY 523 GPKGAAGPPGKAGRGVPPGPPGA-----VGPAGKDGAGAGQPPGPPGAGP---AGER 570

DB 840 GPIGVGAPGPKGPRGAGPPGATGFGAAGRVGPPGPPGAGPPGPPGPPGPPGPPGPP 899

QY 571 GEQGPAGSPGQGLPGPAGPPGEGAGKEGQGVPGDILGAPGSPGAGFPPTGLPGPPGR 630

DB 900 GETGAPRGPEVPPGPPGPPGAGKSPGADGAPGAGSPGTPGPGQIAGQRGVVLPGQR 959

QY 631 GPGSRGPPGADGAVGPKGPPGAGRGSPGPPAGP---KSGPCEAGRPCEAGLPGAKGLTSP 687

DB 960 GFFGLPSPGEPGKQGSSSGRRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1019

QY 688 GSPGPDGKTGPPGAGQDGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 746

DB 1020 GAKGDRGETGAPGPPGAPGAPGPPGAPGKNGDRGETGAPGAPGPIGAPGARGPAGPQ 1079

QY 747 -----PGAVGPAKDGAGAGGPPGPPGAPGAGRGEGQPPGAGSP 783

DB 1080 GPRGDKGEGEQDGRGKIHGRGSGGQPPGSPSGEGQPPGASGAPGPPGPPGAGSP 1139

QY 784 ---GFQGLPGPAGPPGEGAGKPGEGQVPGDILGAPGSPGAPG 820

DB 1140 GKDLNGLPGPIGPPGPPGRTGSDGAPGPPGPPGPPGPPG 1179

RESULT 3

CGCHIS

collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)

C;Species: Gallus gallus (chicken)

C;Date: 12-Aug-1981 #sequence revision 06-Jul-1982 #text_change 31-Mar-2000

C;Accession: A90458; A90181; A02857

R;Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.

Biochemistry 21, 2048-2055, 1982

A;Title: Amino acid sequence of chick skin collagen alpha1(I)-C88 and the complete prima

A;Reference number: A90458; PMID:82231995; PMID:7093229

A;Accession: A90458

A;Molecule type: protein

A;Residues: 1-1036 <HIG>

A;Experimental source: skin

A;Note: this is the latest in a series of papers from these workers elucidating the sequ

R;Eyre, D.R.; Glimcher, M.J.

Biochem. Biophys. Res. Commun. 48, 720-726, 1972

A;Title: Evidence for a previously undetected sequence at the carboxyterminus of the alp

A;Reference number: A90181; PMID:72243016; PMID:5047697

A;Accession: A90181

A;Molecule type: protein

A;Residues: 1037-1042 <EYR>

A;Experimental source: skin

A;Note: residues 1037-1042 above correspond to the carboxyl end of the protein

C;Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some

C;Comment: of the positions at the third position of the tripeptide repeating unit (

C;Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in po

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyrrolutamic acid; trimer;

F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 57.7%; Score 2679; DB 1; Length 1042;

Best Local Similarity 53.7%; Pred. No. 1.8e-126;

Matches 540; Conservative 30; Mismatches 250; Indels 186; Gaps 16;

QY 1 GPPGPGPTGLPGPPGRRGGPGSGRFPAGDGVAGPKGA-----GER 42

DB 23 GPRGLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 82

QY 43 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 93

DB 83 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 142

QY 94 -----GRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 141

DB 143 GLPGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 202

QY 142 GKDGAAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 186

DB 203 GPGSRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 262

QY 187 GPKGEGVPGDILGAP-----GFSGA-----GPPGPTGLPGPP 219

DB 263 GAPGPKNGSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 322

QY 220 GERGPPSGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 279

DB 323 GERGAPSGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 382

QY 280 GSPGPDGKTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 339

383 GSFGPDGKTGPPGPAQDGRPGAPPPGARGAGVWGFPGPKGAAGEPGKGERGAPDPP 442

340 GAVGPAGKDGAGAGQCPGPAAGPAGERGEQGFAGSGFQGLFGPAGPFGAEAGKPGEQGV 399

443 GAVGAAGKDGAGAGQCPGTPGACERGEQGFAGAPFGQGLFGPAGPFGAEAGKPGEQGV 502

400 GDLGAPGSGPAGE-----PGP-----TGLPGPGGERGPP 429

503 GNAGAPGFGAGARGERGFGERGVQPPPGQPRGANGAFGNDGAKGDAGAPAGNEGPP 562

430 GSRGPPGADGVAGPKGPAGERGSPG-----AGPKSGFCEAGRP---GGA 471

563 GLEGMEGERGAAGLPCAKGDRGDPGFKGADGAPGDKGLRGLTGPICTPPGAPAGDCKGEA 622

472 GLPGAAGLTCGSGSPGDPDKTTPPGFA-----GQDGRGPP 507

623 GPPGPAAGPTGARGAPGDRGEPGPPGAPGAGPGADGQPGAKGTGDKAGKDGADGPPGA 682

508 GPPGARGQAGVMGFPKPKGAAGEPGKAGRGVFPGPVAGVPAGKDGEGAGAQGPPGAPGA 567

683 GFTGAPGPAAGZVGAAGPKGARGAGSPGATGTFGAAGRVGPPGPGSNIGLPGPPGPAKZ 742

568 GE---RGEQGPAGSGFQGLPGAPGPPGAEAGKPGEQGVPGDLGAPGSPGAGEPGTGLP 624

743 GSKGPRGETGPAAGPPEGPAAGPPGPPGSGKSGFADGPIGAPGTPGQGIAGQGVGLP 802

625 GPPGERGGPSRGFPADGVAGPKGPAGERGSPGAPG---KGSPGEAGRPGEGALPGAK 681

803 GQRGERGFPLPGSPGEPCKQSPGSGASGERGPPGPMGPPGLAGPPGEGAREGAPGAEGAP 862

682 GLTGSFGSGPDGKTTPPGPAQDGRPGPPGPGAGGQAGVMGFPKPKGAAGEPGKAGER 741

863 GRDGAAGPKDGRGETGAPGPPGAPGAPGPPVGPAGKNGDRGETGAPGAPGPPGAGAR 922

742 GVFGPPGAVGPAKGDGE-----AGAQQGP-----GPAQAGER 774

923 GPAGPQPRGDKEGTGEQGRGMKGRHGFSGLQCPFPFGAPGEGEQGFGSGASGAPGPRGPP 982

775 GEQGPAGSGFQGLPGAPGPPGAGKPGEGEQGVPGDLGAPGSPGAG 820

983 GSAGAAGKQGLNGLPGIPGPPGRGTGTVGVPVGPVPPGPPGPPG 1028

RESULT 4

T45467

collagen alpha 1(II) chain precursor [imported] - horse

N;Alternate names: type II collagen

C;Species: Equus caballus (domestic horse)

C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004

C;Accession: T45467

R;Richardson, D.W.; Dodge, G.R.

submitted to the EMBL Data Library, June 1996

A;Description: Cloning of equine type II collagen and modulation of its expression in

A;Reference number: Z2977

A;Accession: T45467

A;Status: preliminary; translated from GB/EMBL/DBD

A;Molecule type: mRNA

A;Residues: 1-1418 <R>

A;Cross-references: UNIPROT:Q28396; EMBL:U63258; PIDN:AAB05773.1

A;Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology

Query Match	53.4%	Score	2477	DB	2	Length	1418
Best Local Similarity	50.6%	Pred. NC	2.4e-116				
Matches	500	Conservative	48	Mismatches	272	Indels	168
Gaps							17

Qy	1	GPPGEPTGLGPPGRRGGSGRFGAAGVAGKCPAGRGSPGAGPKGSPGNGRP	60
Db	156	GNPGEPPGVSGMGPGRPPGPGKPCDDGEAKPKSGRGPPQGGARGPPTGLP	215
Qy	61	-----GGAGLFGAKGLTGSP-----GSPGP-----DGKTPPPQPA---	90
Db	216	GVKGHRGVPGLDGAKGACAGPVKGSGSGPCNGSGPGPMGRGLFGERGRTGAGAAAG	275

91 QY GDGCRPGPPGPGARGQAGVGMFFGPKGAACGPKKAGER-----GVDPG- 134
 276 DDB GNDQPGAGPPGVPVFGAGGPFPGAACAKEAGFTGARGPEGAQGPGRGPGCTPGSPGPA 335
 135 QY -----PGAVPGAGKDGACAGAGPPPGAPAGAGRGEGQP-----AGSPGFQGLP 177
 336 DDB GAAGNPGTDGIFCAKGSAGAFIAGAPFGPPGPPGQATGLPGRXGTGEPGIAGFK 395
 178 QY GPAGPPGAGKPGCGQVPCDILGAPCSPAGEPGPTGLPGPPGRRGGPGSGFGPCADGVA 237
 336 DDB GEQPGKPGFPAGPGAGPPGAGEKRGARGEPGAGVPVPPGGERGAFGNGRGGFPQDGLA 455
 238 QY GPKGAPAGRRGSPGAPGKSPGEAGRPGEAGLPGAKGLTGSPGSPGPGTKTTPPGPAGQD 297
 456 DDB GPKGAPGRRGSPGLAGPKANGADPCRPGCEPLPGARGLTGRPGDAGPQCKVGPSPGAPGD 515
 238 QY GRPDPGPPGARGQAGVGMFFGPKGAACGSPKAGRGVPPGAGVGPAGKDGCEAGACQPP 357
 516 DDB GRPDPGPPGARGQAGVGMFFGPKGANGEPGKAGEKGLPGAPLGRGLPKDGETGAAGPP 575
 358 QY GPAGPAGRRGEOGPAGSPGFQGLPGPAGPPGAEAGKPGEGQGVPGDILGAPGSPGAGE-- 414
 576 DDB GPAGPAGRRGEOGAPGSPGFQGLPGPPGPGEGGKPGDQGVPGGEAGAPGLVPRGERGFP 635
 415 QY GPTGLP-----GPPGRRGSGRRGPPCADGVAGPKGPA 447
 636 DDB GERGSPGAGLQAGRLPGTPTGDFPKASGPAGPPGAQGPFGLOQMPGERGAAGIAGPK 695
 448 QY GERGSPGAPGKSGPGE-----AGRPGEAGLPGAKGLTGSPGSPGPD 489
 696 DDB GRGDVGEKPGKAGKDCGRGLTGPIPPGPAGANGKEKGVGPPGACTAGARGAPGR 755
 490 QY GXTGPPGAGQDGRGPPPPGAPAR-----GQAVGMFFGPKGAAGEP-----GKA 534
 756 DDB GETGPPGAGFAGPPGADQOPAKGEOGKAGDAGFPQGPQSPGAPGQPGTGVTPK 815
 535 QY GERGVPPGPGA-----VCPAGKDGAGAQGPPGPA--GPACERGECQGPAGSPGFQ 582
 816 DDB GARGAGQPPGATGPGCAARVPPGPNMGVPPGPPGPKGDKGPKGARDSDGPPRAGDP 875
 583 QY GLPGAGPPGAEAGKPGEQCVPGDLAGPSSGAPGPGTGLPFPGERGSGPSRGFPFGAD 642
 876 DDB GLQGPAGPFGKEGPPDGDGSPDPGPPGQLAGRGIVGLPQQRGERGFPGLGPPSGEP 935
 643 QY GVAGPKGAPAGERGSPGAPGKSPGAEAGRPGEAGLPGA-----KGLT 684
 936 DDB GKQAGPAGSGDRGPPGPPVPPGLTGPAGPFRGRTGADGPPGDRDGAAGVKGDRGAGAL 995
 685 QY GSPGSPGPKTKTTPPGPAGQD--GRPPGPPGPPGARGQAGVGMFFPKGAAGPFGKAGE- 740
 996 DDB GAFCAFGPPGSPGCPAGTGTQCDRGEAGAQGWPAGPAGARGLPGPQGPGRGKGEAGEA 1055
 741 QY -----RGVPPGPPGAVGPAKDKGEAGAGQPPGPPGAPAGERGECQGPAGSPFGQLPGPA 792
 1056 DDB GERGLKHGRGFTGLQGLPPPGSPGQASGAPGSPGPRGPPGVGSPGKDGANGIPGI 1115
 793 QY GPPGKAGKPGECQGVPGDLAGPSPGAG 820
 1116 DDB GPPGPRGRSGETGAPGPPGPNPDPGPPG 1143

RESULT 5
 CGHU6C
 collagen alpha 1(II) chain precursor [validated] - human
 N;Alternate names: procollagen alpha 1(II) chain
 N;Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen
 C;Species: Homo sapiens (man)
 C;Date: 28-May-1986 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
 C;Accession: A38133; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S6
 7250; I37251; I37252; I37253; I37254; I55338; I59535; I61910
 R;Ryan, M.C.; Sieraski, M.; Sandell, L.J.
 Genomics 8, 41-48, 1990
 A;Title: The human type II procollagen gene. identification of an additional protein-co-

RESULT 5
CGH06C
collagen alpha 1(II) chain precursor [validated] - human
N:Alternate names: procollagen alpha 1(II) chain
N:Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen
C:Species: Homo sapiens (man)
C:Date: 28-May-1986 #sequence revision 01-Sep-1995 #text change 09-Jul-2004
C:Accession: A38123; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64574; S6
7250; I37251; I37252; I37253; I37254; I55338; I59335; I61910
R:Ryan, M.C.; Sieraski, M.; Sandell, L.J.
Genomics 8, 41-48, 1990
A:Title: The human type II procollagen gene: identification of an additional protein-co

DB	535	GP	KAGPGRG	PSGLIAG	PKGANGDPR	GRPEGLPCARGLTGRPDAGPQGKVGPSGAPGED	584		
QY	298	GR	POPPG	PGFARGOAGV	VMGPPGPKGAAGE	FGKAGERGVPPGPAVGPAGKDXGAGAQGPP	357		
DB	585	GR	PPPG	PGQARGOAGV	VMGPPGPKGANGEP	FGKAGEKGLPGAPGLRGLPGKDXGTGAEGPP	644		
QY	358	GP	APG	AGERG	QGFAGSPGF	QGLPGPAGPGEAGKPGEQCVGPDGLGAPGSPSGPAGE---	413		
DB	645	GP	APG	AGERG	QGFAGSPGF	QGLPGPAGPGEAGKPGEQCVGPEAGAPGLVGRGERGFP	704		
QY	414	---	---	---	---	---	---		
DB	705	GER	SGP	GAQGLQGP	RGGLPGTGTDP	GKASGPAGPQAQGLQGMGPRGERGAAGIAGPK	764		
QY	448	GER	SG	---	---	---	---		
DB	765	GD	RGDVGEK	GP	EGAPGKDGCRGLT	GTGPPGACANGKEGVEVGPAGSAGAGAPGER	824		
QY	484	GS	PGD	GGTGTG	PPAGDGGRP	PGPPGARGAQAGVMGFP	GPKAAAGEP-----GKA	534	
DB	825	GE	TPG	TSGTAG	PPGADGGQ	GKAGEGEAGQKDAGAFQCPQSPS	GAPGQGGTGVGPK	884	
QY	535	GER	GV	PGPPGA	-----	VGPAKGGEAGAGPPGPA	---GPAGERGEQGPAGSPGFQ	582	
DB	885	CARGA	QGP	PGATGFP	FGAAGRVGPP	SGNSGNFGPPGPPGSGKDGPKGARGDSGPPGRAGEP	944		
QY	583	GL	PPAG	PPG	BAGKPG	EQGVGPDGLGAPGSPGAGEP	QGTGLPGPPGERGCGPSRGSPFGAD	642	
DB	945	GL	Q	PAG	PPG	KEGEBDDGSGAEGPPQGLAGQ	RGIVCLPQQRGERGFPGLPGSPGE	1004	
QY	643	GV	AGK	GPAG	RGSPG	PAGPKSGPGEAGRGEAGLPGA	-----KGLT	684	
DB	1005	QOQ	GAP	GASGRG	PPGPPGVP	PLGTGPAGEPGRSGSPGADPPORDG	GAAGVKGBRGETGAV	1064	
QY	685	GS	PSG	PGD	DKTGP	PPG-----	AGQDGRPGPPPGQARGAQGVMGFP	GKGAAGEP	735
DB	1065	GAP	AP	GP	PGSGPPG	PAGPTGKQDRGEAGAGP	MPMGSPGAGARGITQPPQGP	PRGDKGAGEP	1124
QY	736	GK	---	AG	RGV	PPGPAVG	PAGKDXGAGAQCPGPPAGPAGERGE	QGPAGSPGFQGLPGPA	792
DB	1125	GE	RLK	HRGTG	TGLOGL	PPGPPGSGD	OGASGAPGSPRPPFPV	PSGKDXGANGIFGI	1184
QY	793	GP	PEAGK	PG	EQGV	PGDGLGAP	PGSGGPAG	820	
DB	1185	GP	PPR	GRSG	SETG	PAGPPG	PPGPPGPPG	1212	

RESULT 6

A:1182

collagen alpha 1(II) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C>Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999

C:Accession: A41182; A44885

R:Metasaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

J. Biol. Chem. 266, 16862-16869, 1991

A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and

A:Reference number: A41182; MUID:91359489; PMID:1885613

A:Accession: A41182

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-1419 <MET>

A:Cross-references: GS:M65161

R:Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.

Development 111, 945-953, 1991

A:Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartila-

A:Reference number: A44895; MUID:91347939; PMID:1879363

A:Accession: A44885

A:Molecule type: DNA

A:Residues: 1-28 <HE>

A:Cross-references: GS:S63180; NID:G234368; PID:AAB13627.1; PID:G234369

A:Note: sequence extracted from NCBI backbone (NCBIN:G31390, NCBP:G31392)

C:Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology

885 GARGAQPPGNTGTPFGAAGRVGPPGANGNPGPAGPPGAGKDGKGVYRSDSGPPGRAGDP 944
583 GLPGPAGPPGAGKXPGQGVDPDLGAPGSPGAPGEPGTLGLPDPGRRGPGGSRGPPGAD 642
945 GLQGPAGAPGKGBGDDGSLDPPGPPQGLAQGRGIVGLPGQGRGSPGGLPGSPGEP 1004
643 GVAGPKPAGERSGSPGAPGKSGPGEAGRPGEAGLPGA-----KGLT 684
1005 GKQAFAGSRRGPPGVPPGTLTPAGEPRGSPGADGPPGRDGAAGVKDGEGETAL 1064
685 GSPGSPGDPGKTGPPGP-----AGQDRGPPGPPGARGQAQGVMGPPGPKGAAGEP 735
1065 GAPCAPGPPGSPGAPGTPGKQDRGEAGAQGMGSPGAPARGIAGPQGRGKGSSEGEQ 1124
736 GK---AGERGVPGPAGVAPGKDGEGAGAQGPPGAPGAPGRRGQGPAGSPGGLPGPA 792
1125 GERGLKGRHTGLQLGLPDPGSDQAGAPGSPGPRGPPGVGSGKDGSGNGIFGPI 1184
793 GPPGEAGKPGEQGVPGDLGAPGSPGAPG 820
1185 GPPGRRGSGETGVPVGGSPGSPGPPGP 1212
RESULT 8
collagen alpha 1(II) chain precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: B40333
J. Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A:Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A:Reference number: A40333; MUID:92011898; PMID:1918153
A:Accession: B40333
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-1486 <SUA>
A:Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M63595
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:37-96/Domain: von Willebrand factor type C repeat homology <WVC>
F:1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Query Match 51.7%; Score 2401; DB 1; Length 1486;
Best Local Similarity 48.9%; Pred. No. 1.4e-112;
Matches 493; Conservative 48; Mismatches 276; Indels 192; Gaps 16;
QY 1 GPPGPPGTLGPPGRRGSGRFPQADGVAGPKGP-----AGER 42
Db 209 GRRGPPGTLGAPGQGFQGNPGEPPGAGGPMGRRGPGSGKXFGDDGEAGKPKSGER 268
QY 43 GSPGAPGKSGPGEAGRP-----GEAGLPGAK---GLTSGSPGSPG----- 80
Db 269 GPPGQAGRFPGTGLPGVKGHRGYPGLDGAKGEAGAGAKGEGGATGEAGSPGPMGPR 328
QY 81 -----DGKTGPPGAGQDGRGPPGPPGARGQAGVGMFPCKP 117
Db 329 GLPGERGPPGSGAAGARNGLGPPAGPPGPPVAGAPFPFGAPGSKGEGAGTGAARPE 388
QY 118 GAAGEPKAGERGVPGPAGVAGPAGKDGAGAGQGPFPAGPAGRGEGGCPAGSPGFGQLP 177
Db 389 GAQGPRESGTPGSPGAPAGSNFGPTDIPGAKSGSGGFIAGAPGFPGRGPPGPPQAT 448
QY 178 GPAGPPGEAGKP-----GEQVFDLGAAPRS---GPA-----GEPTGLPGPP 219
Db 449 GLPLGPKGTGDPGAVGKEQGPKEGIGSAGPQAGPAGBEGKKGARGEGAAAGNPP 508
QY 220 GERGGPGSGRFPAGDGVAGPKGAPGRRGSPGAPGKSGPGEAGRPGEAGLPGAKGLTGP 279
Db 509 GERKAFNGRFPQDGLAGPKAPGERGVPLGLGGKGGNDGPRGEPGLPGARGLTGRP 568
QY 280 GSPGPPGKTPPGPAGQDGRGPPGPPGARGQAQGVMGFFGPPKGAAGEPKAGERGVP 339

569 GDAGPQGVGSPGASGEDRPPGPPQAGRGQGVYMGFPGPKGANGEPCKAGEKGLVAP 628
340 GAVGPAGKDXGAGAGPAGPAGPAGRGEGEQGAPGSPGFGQLPGPAGPPGEGAKPGEQGV 399
629 GURGPFKKDGETGSGQPNPAGPAGRGEGEQGPPGSPGFGQLPGPAGPPGEGAKPGDQGV 688
400 GDLGAPGSPGAGE---PGFTGLPDPGERGPGGRRGPGADGVAGPKGPPAGERSGPPA 456
689 GEAGAPGLVGRGERGFPGERGSGPQGLQGRPLGFTPTDGPKGASGSPGNAQGP 748
457 GPKGSPGEAGRGEGAGLPGAKGLTSGPGSPGDPGTPGPPAGQDGRGPPGPPGARGQA 516
749 GLQGMPPGERGAAGISGPKDRGDTGKGPGEAGKDGSRGLTGPFGPAGPAGNKEKES 808
517 GVMGPPGPPGAAGEPKAGERGVPGPAGVAGPAGKD-----GEAGAQGPP 561
809 GPGSPGPIVGARGAPCDRGENGPPGPPAGPAGPAGADGQSGKDGQESQKGDAGAPGQ 868
562 GPA-----GPAGRGEGQAGSPGFGQL-----PGPAGPPGEGAKP 597
869 GPGGAPGQGTGVPFGPKGARGAQGAPAGATGFPGAAGRVGTGPNNGNPPGPPGPGSAGK 928
598 GEOGVPGDLGAP-----GPSGPAGEPFTGL----- 623
929 GPKSVGDDAGPPCRAGDPGLQGAAGAPGKGBPGEDGSPGDPGPPGQGLSGQRGIVGLP 988
624 -----PGPPGERGPPGSRGPPGADGVAGPKGAPAGERGSPGAPGKSPGEGAPGEB- 674
989 GQRGERRFPCLPGSPGEPGKQSGSGRGPFPV---GPPGLTGPSPGEPGREGNPGSD 1045
675 -----AGLPGAKGLTGPSPGPPGDKTGPAGPAGQDGRGPPGPP 714
1046 GPPRGDAGTIKDRGETGFLNAPGAPGAPGAFGSGVGTGKQDGRGSPGQGLPGSGPA 1105
715 GARGQAGVMGFPFGKGAAGEPPKAGB---RGVPPGPPGAVGPPAGKGEAGAQGPPGAPGA 771
1106 GARGLAGPQPRGDKGEAGEAGERGQKGRGTGLQGLPQPPGSGADQDAGTGPAGPGR 1165
772 GERGEOGAPGSPGFGQLPGPAGPPGAGKGEQGVPGDLGAPGSPGAPG 820
1166 GPPGPPGVPSPGKDSNGISGPIGPPGRGSGETGTPSGPPGQPPGPPG 1214
RESULT 9
A40333
collagen alpha 1(II) chain precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
C:Accession: A40333
J. Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A:Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis e
A:Reference number: A40333; MUID:92011898; PMID:1918153
A:Accession: A40333
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1492 <SUA>
A:Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M63596
A:Note: this sequence is presented as substitutions relative to another sequence in a f
es they replace; the appropriate interpretation of the sequence figure was reconstructe
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:37-96/Domain: von Willebrand factor type C repeat homology <WVC>
F:1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Query Match 51.5%; Score 2389; DB 2; Length 1492;
Best Local Similarity 50.6%; Pred. No. 5.7e-112;
Matches 477; Conservative 59; Mismatches 275; Indels 132; Gaps 14;
QY 1 GPPGEPGPTGLGPPGRRGPPGRRGFPFAD-----GVAGPKGAPGRRGPPGAPGPK 51
Db 272 GPPGQAGRFPFGTGLPGVKGHRGYPGLDGKGEAGAGAKGSGASGEAGAPGMPGR 331

QY 52 GSPGEAGRCEAGLPGKAGLGTGPGSGPGDGKGTGPPGAGDGRPGPPGPPGARGAGQVM 111
 Db 332 GLPGERGRSSGAAGAR---GNDGLPGP---AGPPGVGPAGARFGFGAAGSGKEAGPT 385
 QY 112 GPFGPKGAAGGPKAGRGVPGP-----PGAVGPKAGDGEAGAAGPPGPPAGPA 159
 Db 386 GARGPEGAQGPGRGSGTSPGSPGSGASGNPGTGTGIPGAKSGSGASGIAGAPGFPGRGPP 445
 QY 160 GERGEQGP-----AGSPGFQGLPGPAGPPGAGRGPGQVPGDGLGAPGSPGAPGPT 213
 Db 446 GPGQATGLPKGTGTDPGIAFGKEHGPKGISAGPQAGPAGPAGGEGKGAERGEGAA 505
 QY 214 GLPFGPGERGSGRSGFGPGADGVAQPKGPAGRGSPGPGAGPKSGPGEAGRGAGLPGAK 273
 Db 506 GPLGPPGERGAPGNRGFGPQDGLAGPKGAPGREGVPLGLGPKGNGDPRGPGELPGAR 565
 QY 274 GLTSGSPGPPGKGTGPPGAGDGRPGPPGPPGARGQAGVMPGPPGKGAAGRGKAGER 333
 Db 566 GLTGAPGADGPPGKGVGSPGAAGEDGRPPGPGQARGQPGVMPGPPGKGAAGRGKAGEK 625
 QY 334 GVPFPFGAVGPKAGDGEAGQAGPPGAPGAGRGQGPAGSPFGQGLPGPAGPPGEGAKP 393
 Db 626 GLGAPGLRGLPKDGETGAQGNPGAPGAGRGEGQPPGSGFQGLGPPGSGEGGKP 685
 QY 394 GEQGVPGDLGAPGSPGAGE---PPTGLPDPGRRGPGSGRFGPGADGVAGPKGAPAGER 450
 Db 686 GDQGVPGGAGAGLVPRGERGFPGERGSGPQGLQSGRGLPGTGTDPKAGATGSPGPN 745
 QY 451 GSPGPGAGPKGSPCEAGRGCEAGLPGKAGLGTGSPSGPDGKTPGPGAGDGRPGPPGPP 510
 Db 746 GAQGPGLQGNPGERGAAGISGPKDRDGTGKGPGEAGPKDGRGLTGLPPEPAGPN 805
 QY 511 GARGQAGVMPGPPGKGAAGEPKAGRGVPPGPPGAVGPAKD-----GEA 555
 Db 806 GEKSGSGSPGPGIVGARGAPCDRGENGPPGPPGAPGPPGSDGQAGLKGQSGSQKXDA 865
 QY 556 GAQGPFGPA-----GPAGERGQGPAGSPGFQGL-----PGPAGPP 591
 Db 866 GAGPQGSAGPAGPQGTGTVNGPKARGAQAGATGFPGAAGRVGTGNGNPPGPPGPP 925
 QY 592 GEAGKPGQGVPGDLGAPGSPGAGEPGFTGLPFPGERGPGSGRFGPGADGVAGPKGPA 651
 Db 926 SSAGKEGPKVGRD---AGPTGRAGDPLQGPAGAPGKGEPEGEDGSGDPGSGPQGLS 982
 QY 652 GERSPGAPGKGS-----PGEAGRGAEGLPKAGLGTSPGSPGDPGKTPGPPGAGQ 705
 Db 983 GNRGIVLPGQRGERGFPGLPSPGSEPKQGGSGSDGRPPGPPGVLGTGAGETGRE 1042
 QY 706 GRPFPFPFGARGQAGVMPGPPGKGAAGEPKKAGRGVPPGPGAVGPAKDGAGAGQGP 765
 Db 1043 GNPSSDGPFGDGTATGKIDRGETGPIGAPCAPSPGAPGVGTGTGQDRGSGSGPL 1102
 QY 766 GPAGP-----AGRGEGQGPAGSPGFQGLP-----SGPAG 789
 Db 1103 GPSGAPGARGLNGPQGRGKGEAGEAGRGKQHRGFTGLQGLPFPPTAGDQAGTGA 1162
 QY 790 -----GPAGPPGAGKFGEGVPGDLGAPGP-----SGPAG 820
 Db 1163 GPSGPRGPPGVPVSGKDGSGNIGFPIGPPGPRGRSETGPAG 1205
 RESULT 10
 CGB07S
 collagen alpha 1(III) chain - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 04-Dec-1996 Sequence revision 04-Dec-1986 #text change 09-Jul-2004
 C:Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946
 R:Fietzek, P.P.; Allmann, H.; Kauterberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979
 A:Title: The covalent structure of calf skin type III collagen. I. The amino acid sequence
 A:Reference number: A02862; MUID:80026026; PMID:488906
 A:Accession: A02862
 A:Molecule type: protein

A:Residues: 1-242 <FIE>
 A:Cross-references: UNIPROT:P04258
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979
 A:Title: The covalent structure of calf skin type III collagen. II. The amino acid sequence
 A:Reference number: A38001; MUID:80026027; PMID:488907
 A:Accession: A38001
 A:Molecule type: protein
 A:Residues: 243-422 <DEW1>
 R:Bentz, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979
 A:Title: The covalent structure of calf skin type III collagen. III. The amino acid sequence
 A:Reference number: A38002; MUID:80026028; PMID:488908
 A:Accession: A38002
 A:Molecule type: protein
 A:Residues: 423-571 <BEN>
 R:Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979
 A:Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequence
 A:Reference number: A38003; MUID:80026029; PMID:488909
 A:Accession: A38003
 A:Molecule type: protein
 A:Residues: 572-808 <LAN>
 R:Dewes, H.; Fietzek, P.P.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979
 A:Title: The covalent structure of calf skin type III collagen. V. The amino acid sequence
 A:Reference number: A38004; MUID:80026030; PMID:488910
 A:Accession: A38004
 A:Molecule type: protein
 A:Residues: 809-947 <DEW2>
 R:Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.
 Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979
 A:Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequence
 A:Reference number: A38005; MUID:80026031; PMID:488911
 A:Accession: A38005
 A:Molecule type: protein
 A:Residues: 948-1049 <ALL>
 A:Experimental source: skin
 R:Henkel, W.
 Biochem. J. 318, 497-503, 1996
 A:Title: Cross-link analysis of the C-telopeptide domain from type III collagen.
 A:Reference number: S71946; MUID:96404897; PMID:8803038
 A:Accession: S71946
 A:Molecule type: protein
 A:Residues: 87-106;1017-1029;1037-1049 <HEN>
 C:Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are highly conserved in the type III collagen molecule. The type III collagen molecule is a trimer of identical chains, linked to each other by interchain disulfide bonds. The type III collagen molecule is a trimer of identical chains, linked to each other by interchain disulfide bonds. The type III collagen molecule is a trimer of identical chains, linked to each other by interchain disulfide bonds.
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline; F:1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>
 F:1-14/Region: amino-terminal nonhelical telopeptide
 F:15-1040/Region: helical
 F:587-589/Region: cell attachment (R-G-D) motif
 F:752-754/Region: cell attachment (R-G-D) motif
 F:875-877/Region: cell attachment (R-G-D) motif
 F:878-880/Region: cell attachment (R-G-D) motif
 F:935-937/Region: cell attachment (R-G-D) motif
 F:1041-1049/Region: carboxyl-terminal nonhelical telopeptide
 F:105-107,109,938,950/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:107,950/Modified site: allysine (Lys) #status predicted
 F:107/Binding site: carbohydrate (Lys) (covalent) #status experimental
 F:1040,1041/Disulfide bonds: interchain #status predicted

Query Match 51.3%; Score 2379; DB 1; Length 1049;
 Best Local Similarity 49.8%; Pred. No. 1.4e-111;
 Matches 479; Conservative 48; Mismatches 250; Indels 144; Gaps 18;
 QY 1 GPPGEPGTGTPGPPGRRGGGSGRGF-----PGADGVAGPKGAPAGER 42
 Db 78 GRPGRGPRGPPGPPGPKMGPGAGMGFGPMKGRHGRGFDGRNGEKGPAGLKGNGVPGED 137
 QY 43 GSPGAPGKSGPAGRGPEAGLPGAKLGTGSPGSPGDKTGPDPGAGQDGRPPGPP 102

Db 138 GAGPMPGRCAPGERGSP---GLPGAAGARGNDGARGSDGQGPFGPPTAGTGFPGSPGAK 194
Qy 103 GARGQAGVMGFPFGKGAAGSPGKAGRGVFPGP-----GAVGPA---GKQGEA 147
Db 195 GEVGPAGSPGSSGAPGQRPQGHAGAPFGPPGPGSDGSPGKGMGPAGIPGAPGLI 254
Qy 148 GAQGPFGAPGAGERCEQAGSPGFGQLPGPAGPPEAGKPGQGVPGDGLGAPGSPGA 207
Db 255 GAGSPGPPGTVGVRGAGAGPFGKXGAKGDPFRGERGEAGSPGIAAGPKGEDKQGSP 314
Qy 208 GEPGPTGLPQPPGERGSPGSPGAGPAGKAGPAGERGSPPGAPGKSPGAGRPGEA 267
Db 315 GEPGAGNLPAGAGERGVPFRGAGANGLPFGKGGPPGDRGGGPGAPRGVAGEPGRNGLP 374
Qy 268 GLPKAGLITGSPGSPGDPGKTGPPGAGODGRFPFGPPGARGQAGVMGFPFGKGAAGCP 327
Db 375 GPGQLRGI PGPSPGSPGNSKFPFGPGSQGTGRFPFGSPGPRGQPGVWGFPGKNDGAP 434
Qy 328 GKAGRGVFPFGAVGAPGAKDGAQAGPFGPAGPAGERGEQGPAGSPFGQLPGPAGPP 387
Db 435 GKNGERGPGGPGFPGPAGKNGETGPGPPGTPGSGDKGDTGPPGQGLQGLPGLTSGPP 494
Qy 388 GEAGKPGQGVPGDLGAP-----GPSGAPGEGTGLPFGPGERGGSGRFGCADGVA 441
Db 495 GENGKPGQEPGKGEAGAPGPGKGDSPGAPGERGPPGAGPPGPGGAGAPPGPBGKGA 554
Qy 442 GPKGAPGERGSP---GPAKPGSPGAGPAGPAGLPGKAGLITGSPGSPGDPGKT---GPP 495
Db 555 GPPGSPGASGTPGLQMPGEPGPGFGPKGKSPGSGVDPGAPKDGKPGGTPGIPGP 614
Qy 496 GPAGQDGR-----PGPFGPPGARGO---AGVMGFPFGKGAAGPFGKAGRGVP 540
Db 615 GPAGQPDGKXGSPAGVPFGIAGPRGFGPGERGQGPFGAGFPAGQNGEPGAKGERGAP 674
Qy 541 -----GPPGAVGAPKDGGAAGQGP-----GPAGPAGERGEQ 573
Db 675 GERKEGPGCAAGPAGGSPGAPGPGQGVKGRGSPGPGGAAGFPGGRGPPGPPGSGNPN 734
Qy 574 GPAGSPGFGQLPGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 623
Db 735 GPPGSSGAPKQGPFGPPGSPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 794
Qy 624 -----PGPFGPGRGSPGFPAGDGVAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 666
Db 795 GLTARGLAGLPGMPGARGSPGQGIKNGKSPGPGSGQNGERGPPGQGLPGLAGTAGEP 854
Qy 667 GEAGRPGEAGLPGA-----KGLTSPGSPGPDGKTGPPGP-----AGQGRPGPP 711
Db 855 GRDNTGSDGLPDRDAPKAGDRGENGSPGAPGAPGPPGPPGPPGPPGPPGPPGPPGPP 914
Qy 712 GPPGARGQAGVMGFPFGKGAAGPFGKAGRGV-----PGPPGAVGPPAGKDGAGAQ 762
Db 915 GPGAPGAPGAGSRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 974
Qy 763 GPPGAPGAPGAGERCEQAGSPGFGQLPGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 819
Db 975 GSPGAPGAPGVPFGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1034
Qy 820 G 820
Db 1035 G 1035

RESULT 11

S59856
collagen alpha 1(III) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text_change 09-Jul-2004
C;Accession: S59856; S62120; S16373
R;Toman, P. D.; de Crombrughe, B.
Gene 147, 161-168, 1994
A;Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA
A;Reference number: S59856; MUID:95011609; PMID:7926795

A;Accession: S59856
A;Molecule type: DNA
A;Residues: 1-1464 <TOM>
A;Cross-references: UNIPROT:P08121; EMBL:X52046
R;Toman, P.
submitted to the EMBL Data Library, November 1994
A;Reference number: S62120
A;Accession: S62120
A;Molecule type: DNA
A;Residues: 1-866, 'G', 868-1464 <TOM>
A;Cross-references: EMBL:X52046; NID:G575321; PIDN:CAA36279.1; PID:G575322
R;Matsaeranta, M.; Toman, P.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs
A;Reference number: S16176; MUID:91274355; PMID:2054384
A;Accession: S16373
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1442-1464 <MET>
A;Cross-references: EMBL:X57983; NID:G50476; PIDN:CAA41048.1; PID:G50477
C;Genetics:
A;Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 298/3; 315/3; 333/3; 351/3; 369/3; 387/3; 405/3; 423/3; 441/3; 459/3; 477/3; 495/3; 513/3; 531/3; 549/3; 567/3; 585/3; 603/3; 621/3; 639/3; 657/3; 675/3; 693/3; 711/3; 729/3; 747/3; 765/3; 783/3; 801/3; 819/3; 837/3; 855/3; 873/3; 891/3; 909/3; 927/3; 945/3; 963/3; 981/3; 999/3; 1017/3; 1035/3; 1053/3; 1071/3; 1089/3; 1107/3; 1125/3; 1143/3; 1161/3; 1179/3; 1197/3; 1215/3; 1233/3; 1251/3; 1269/3; 1287/3; 1305/3; 1323/3; 1341/3; 1359/3; 1377/3; 1395/3; 1413/3; 1431/3; 1449/3; 1464/3
C;Superfamily: Collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-154/Domain: propeptide #status predicted <PRO>
F;32-92/Domain: von Willebrand factor type C repeat homology <VWC>
F;155-1464/Product: collagen alpha 1(III) chain #status predicted <NAT>
F;1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Query Match 51.2%; Score 2377.5; DB 2; Length 1464;
Best Local Similarity 51.4%; Pred. No. 2.1e-111;
Matches 477; Conservative 56; Mismatches 266; Indels 129; Gaps 17;
Qy 1 GPPGFPPTGLPFPGERGSPGSRGFP-----GADGVAGPAGPAGER 42
Db 233 GPRPGRGRLGPPGKXGAPGMPFGMKHGRGDRNGEKGTGAPGLKXGNGLPDGN 292
Qy 43 GSPGAPGKSPGAGRPGGAGLPGKAGLITGSPGSPGPDGKTGPPGAGQGRP----- 96
Db 293 GAGPMPGARGPGERGSP---GLPGAAGARGNDGARGSDGQGPFGPPTAGTGFPGSPGAK 349
Qy 97 ---GPPGPPGARGQAGVMGFPGPCK---GAAGPAGKAGRGVFPFGVAGVPA---GKQGEA 147
Db 350 GEVGPAGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPG 409
Qy 148 GAQGPFGPAGPAGERGEQAGSPGFGQLPGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 207
Db 410 GARGPFGPAGTNGIPGTRGSPGEPKNGAKGEPGARGERGAGSPGIFGPKGEDKQGSP 469
Qy 208 GPPGPTGLPFPGERGSPGSRGFPAGDGVAGPAGPAGRSPGPPAGPAGSPGAGRPGEA 267
Db 470 GEPGANGLPAGAGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSP 529
Qy 268 GLPGAAGLITGSPGSPGDPGKTGPPGAPGQDORPFPFGPPGARGQAGVMGFPFGKGAAGCP 327
Db 530 GPGFIRGMPGSPGPGFNDKFPFGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPGSP 589
Qy 328 GXAGRGVFPFGPAGVAGPAGKDGAGAQGPFGPAGPAGRGEQGPAGSPGFGQLPGPAGPP 387
Db 590 GXNGERGPGGGLPGPAGKNGETGPPGPPPTGPDGKDGSGPFGPQGLGIPGTGGPP 649
Qy 388 GEAGKPGQGVPGDLGAPGSPGAPGPPPTGLPFPGERGPPGSRGPPGADGVAGPAGKGA 447
Db 650 GENGKPGEPGPKGEVGP-----GAPGKGDSPAGPGERGPPGTAGIFGARGGAGPSPG- 702
Qy 448 GERGSPGAPGKSPGAGRPGGAGLPGKAGLITGSPGSPGPDGKTGPPGAGQDGRPGPP 507
Db 703 ---EGGKPGAGPFGPPGASGSPGLQMPGER---GPGSPGPKGKSPGCGAGAGVPGKD 757
Qy 508 GPPGARGQAGVMGFPFGKGAAGPFGKAGRGVFPFGVAGVAGPAGKDGAGAGAGPAGP- 564

Db 758 GP---RGPAGPIGPPGP---ACQPGDKXGSGSPGLPIAGPGRGGGEGHGPAGP 811
QY 565 -----GPAEREGEQPAGSPFQGLPAGPAGPGE---AGKPGEGVPGDLGAPGSPGA 615
Db 812 GAFQNGEGFCAKGERGAPKGEKGGPFGPAGTGTSGSPAGPQGVKGERGSPGPGTA 871
QY 616 GEGPPTGLPQPPGERGSGRSPFGADGVAGPKGPAGBERGSPGAPKSGSGEAGRPGE- 674
Db 872 GFGGKGLGFPQNNGNPFPSPGAPKGDGPPGAGNSGSPGNGIAGPKGDAGQGEK 931
QY 675 -----AGLCAKGLTSGPSGPGDGKTGP-----FGPAGQDGRP 708
Db 932 GPPGAGQPPGSPGLIAGLTGARGLAGPPGMPGPRGSPGPGIKGESKPGASGNGER 991
QY 709 GPPGP-----PGARGQAGVWGPPGPKGAAGE---FGKAGERGVPPGP 747
Db 992 GPPGPGGLPQGTAGEPRDNGPDSGQPGFGRDGSFGKGRGNGSPGAPCAPGHPGPP 1051
QY 748 GAVGPAGKQGE-----ACAQGPFPAGPAGERGEQPGAGSPGFQGLPGPA 792
Db 1052 GPVPSGKSGDRGETGPAFGSGAPGAPGARGAPGQGRGDKGTERGSGNGIKHRGFP 1111
QY 793 GPPGAGKPGEGVPGDLGAPGSPGAG 820
Db 1112 GNPFPFGSGAAGHQAGTSGPAGPRG 1139
RESULT 12
CGHU7L
collagen alpha 1(III) chain precursor - human
N:Alternate names: procollagen alpha 1(III) chain
C:Species: Homo sapiens (man)
C>Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 09-Jul-2004
C:Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90399
R:Prockop, D.J.
submitted to the EMBL Data Library, February 1989
A:Reference number: S05272
A:Accession: S05272
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1240, 'V', 1242-1466 <PRC>
A:Cross-references: UNIPROT:P02461; EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058
R:Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kulvanliem, H.; Prockop, D.J.
Biochem. J. 260, 509-516, 1989
A>Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human
erences.
A:Reference number: S04642; MUID:89350838; PMID:2764886
A:Accession: S04642
A:Molecule type: mRNA
A:Residues: 1-1196 <ALA>
A:Cross-references: EMBL:X14420; NID:G30057; PIDN:CAA32583.1; PID:G30058
A>Note: the complete sequence is not shown
R:Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
Gene 78, 255-265, 1989
A>Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene
A:Reference number: PE0011; MUID:89378752; PMID:2777083
A:Accession: PE0011
A:Molecule type: DNA
A:Residues: 1-176 <BEN>
A:Cross-references: M26939; NID:G180813; PIDN:AAA52040.1; PID:G180814
R:Tomari, P.D.; Ricca, G.A.; de Crombrughe, B.
Nucleic Acids Res. 16, 7201, 1988
A>Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre
A:Reference number: S01726; MUID:89303360; PMID:3405773
A:Accession: S01726
A:Molecule type: mRNA
A:Residues: 1-170 <TOM>
A:Cross-references: EMBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:G30061
A>Note: the authors translated the codon CAG for residue 154 as His
R:Janeczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 6742, 1989
A>Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
A:Reference number: S04887; MUID:89386015; PMID:2780304

A:Accession: S04887
A:Molecule type: mRNA
A:Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634,
A:Cross-references: EMBL:X15332; NID:G29545; PIDN:CAA33387.1; PID:G930045
A>Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R:Seyer, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A>Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A:Reference number: A90399; MUID:77134724; PMID:557335
A:Accession: A90399
A:Molecule type: protein
A:Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEV1>
A:Experimental source: liver
A>Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact
R:Seyer, J.M.
submitted to the Atlas, December 1977
A:Reference number: A94562
A:Accession: A94562
A:Molecule type: protein
A:Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>
A:Experimental source: liver
A>Note: author submitted corrections to A90399
R:Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A>Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual
splicing.
A:Reference number: I51868; MUID:93304430; PMID:8317500
A:Accession: I51868
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 186-194 <MIL>
A:Cross-references: GB:S62925; NID:G36425; PIDN:AA013937.1; PID:G4261637
R:Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A>Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3
A:Reference number: S59511; MUID:96067614; PMID:7487954
A:Accession: S59511
A:Molecule type: mRNA
A:Residues: 302-423 <CHI>
A:Cross-references: GB:S79877; NID:G1195576; PIDN:AA835615.1; PID:G1195577
R:Seyer, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A>Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr pe
A:Reference number: A90414; MUID:79000343; PMID:687591
A:Accession: A90414
A:Molecule type: protein
A:Residues: 399-675, 'N', 677-727 <SEY3>
A:Experimental source: liver
R:Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A>Title: G to T transversion at position +5 of a splice donor site causes skipping of th
A:Reference number: I55349; MUID:91161621; PMID:1672129
A:Accession: I55349
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 537-605 <LEE>
A:Cross-references: GB:M59312; NID:G180815; PIDN:AAA52041.1; PID:G180816
R:Seyer, J.M.; Mainardi, C.; Kang, A.H.
Biochemistry 19, 1583-1589, 1980
A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from ty
A:Reference number: A90438; MUID:80198282; PMID:6246925
A:Accession: A90438
A:Molecule type: protein
A:Residues: 728-895, 'A', 897-964 <SEY4>
A:Experimental source: liver
R:Lee, B.; Chiodo, A.A.; Lamanche, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Char
J. Biol. Chem. 265, 17070-17077, 1990
A>Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an
A:Reference number: A38303; MUID:91009133; PMID:2145268
A:Accession: A38303
A:Molecule type: mRNA
A:Residues: 861-1015 <COL>
A:Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AA859383.1; PID:G

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 13:23:10 ; Search time 146.729 Seconds
(without alignments)
3219.411 Million cell updates/sec

Title: US-10-658-989A-3

Perfect score: 4640

Sequence: 1 GPGGPGTGLPGPPGRRG.....GEOGVGDLGAPGSPGAGG 821

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2765	59.6	1464	1 CALL_HUMAN	P02452 homo sapien
2	2761	59.5	1461	2 Q6045	Q6045 homo sapien
3	2761	59.5	1464	2 Q8N473	Q8N473 homo sapien
4	2759	59.5	1460	1 CALL_CANFA	Q9xsj7 canis famil
5	2748.5	59.2	1453	2 Q63079	Q63079 rattus norv
6	2732.5	58.9	1453	1 CALL_MOUSE	P11087 mus musculu
7	2732.5	58.9	1453	2 Q810J9	Q810J9 mus musculu
8	2681.5	57.8	1069	2 Q6LAN8	Q6LAN8 homo sapien
9	2679	57.7	1069	2 CAA67261	CAA67261 homo sapi
10	2679	57.7	1453	1 CALL_CHICK	P02457 gallus gall
11	2568	55.3	1445	2 Q93251	Q93251 rana catesb
12	2561.5	55.2	1450	2 Q9YTB4	Q9YTB4 cynops pyr
13	2523	54.5	1449	2 Q802B5	Q802B5 xenopus lae
14	2494	53.8	1225	2 Q6PCL3	Q6PCL3 mus musculu
15	2494	53.8	1225	2 AAH59281	AAH59281 mus muscu
16	2486	53.6	1160	2 Q14046	Q14046 homo sapien
17	2486	53.6	1487	2 Q14047	Q14047 homo sapien
18	2481	53.5	1487	2 Q77753	Q77753 canis famil
19	2479	53.4	1447	2 Q9IB91	Q9IB91 xenopus lae
20	2477	53.4	1418	2 Q28396	Q28396 equus cabal
21	2476	53.4	1418	1 CALL_HUMAN	P02458 homo sapien
22	2475.5	53.4	1420	2 Q90W37	Q90W37 gallus gall
23	2465.5	53.1	1269	2 Q7T227	Q7T227 gallus gall
24	2464	53.1	1442	2 Q62031	Q62031 mus musculu
25	2464	53.1	1442	2 Q62033	Q62033 mus musculu
26	2464	53.1	1459	1 CALL_MOUSE	P28481 mus musculu
27	2464	53.1	1459	2 Q62032	Q62032 mus musculu
28	2455	52.9	1419	2 Q80V13	Q80V13 mus musculu
29	2454	52.9	1419	2 Q63123	Q63123 rattus norv
30	2449	52.8	1419	2 Q80X38	Q80X38 mus musculu
31	2430	52.4	1447	2 Q6P4U1	Q6P4U1 brachydanio

32	2430	52.4	1447	2 AAH63249	AAH63249 brachydan
33	2428.5	52.3	1447	2 Q6U1J5	Q6U1J5 brachydanio
34	2428.5	52.3	1447	2 AAR24536	AAR24536 brachydan
35	2412	52.0	1418	2 Q9W7R9	Q9W7R9 cynops pyr
36	2405	51.8	1492	2 Q6F4Z2	Q6F4Z2 xenopus tro
37	2405	51.8	1492	2 AAH63191	AAH63191 xenopus t
38	2401	51.7	1486	2 Q91717	Q91717 xenopus lae
39	2399	51.7	1486	2 Q7ZTI6	Q7ZTI6 xenopus lae
40	2390	51.5	1449	2 Q6PBI9	Q6PBI9 brachydanio
41	2390	51.5	1449	2 AAH58045	AAH58045 brachydan
42	2386	51.4	1491	2 Q91718	Q91718 xenopus lae
43	2385	51.4	1491	2 Q7ZTM4	Q7ZTM4 xenopus lae
44	2384	51.4	1449	2 Q6NZ15	Q6NZ15 brachydanio
45	2384	51.4	1449	2 AAH66384	AAH66384 brachydan

ALIGNMENTS

RESULT 1

CALL_HUMAN STANDARD; PRT; 1464 AA.
AC P02452; P78441; Q13896; Q13902; Q14037; Q14992; Q15176;
AC Q15201; Q16050; Q7KZ30; Q7KZ34; Q81V15; Q9UML6; Q9UVM7;
DT 21-JUN-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
CN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT SER-1434.
RA Dagleish R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-589 FROM N.A.
RX PubMed=2843432;
RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,
RA Pretorius P.J.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene
RT (COL1A1).";
RN [3]
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockup D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [4]
RP SEQUENCE OF 1-181 FROM N.A.
RX MEDLINE=84270697; PubMed=6462220;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA Myers J., Williams C., Ramirez F.;
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT conservation of a pattern of introns and exons.";
RL Nature 310:337-340(1984).
RN [5]
RP SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.
RX TISSUE=Skin; PubMed=5529814;
RA MEDLINE=71038625; PubMed=5529814;
RA Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 1 and alpha 2 chains of human skin collagen.";
RL Biochemistry 9:4699-4706(1970).
RN [6]
RP SEQUENCE OF 425-1464 FROM N.A.
RX MEDLINE=84080385; PubMed=6689127;

- collagens. A possible role of the carbohydrate in fibril formation.",
J. Biol. Chem. 245:5042-5048(1970).
[16]
SEQUENCE OF 281-302; 402-420; 823-842; 924-944; 1026-1045 AND
1143-1162 FROM N.A.
PubMed=2374517;
Labhard M.E., Hollister D.W.;
"Segmental amplification of the entire helical and telopeptide regions
of the cDNA for human alpha 1 (I) collagen.";
Matrix 10:124-130(1990).
[17]
SEQUENCE OF 425-490; 965-1024; 999-1039 AND 1453-1464 FROM N.A.
MEDLINE=83064528; PubMed=6183642;
Chu M.-L., Myers J.C., Bernard M.P., Ding J.-F., Ramirez F.;
"Cloning and characterization of five overlapping cDNAs specific for
the human pro alpha 1(I) collagen chain.";
Nucleic Acids Res. 10:5925-5934(1982).
[18]
SEQUENCE OF 710-745 FROM N.A., AND VARIANT OI-II ARG-728.
PubMed=2339700;
Wallis G.A., Starman B.J., Zinn A.B., Byers P.H.;
"Variable expression of osteogenesis imperfecta in a nuclear family is
explained by somatic mosaicism for a lethal point mutation in the
alpha 1(I) gene (COL1A1) of type I collagen in a parent.";
Am. J. Hum. Genet. 46:1034-1040(1990).
[19]
SEQUENCE OF 746-781 FROM N.A., AND VARIANT OI-III SER-767.
MEDLINE=95187161; PubMed=7881420;
Forlino A., Zolezzi F., Valli M., Pignatti P.F., Cetta G.,
Brunelli P.C., Mottes M.;
"Severe (type III) osteogenesis imperfecta due to glycine
substitutions in the central domain of the collagen triple helix.";
Hum. Mol. Genet. 3:2201-2206(1994).
[20]
SEQUENCE OF 1187-1220 FROM N.A., AND VARIANT CVS-1195.
PubMed=3170557;
Cohn D.H., Apone S., Eyre D.R., Starman B.J., Andreassen P.,
Charbonneau H., Nicholls A.C., Pope F.M., Byers P.H.;
"Substitution of cysteine for glycine within the carboxyl-terminal
telopeptide of the alpha 1 chain of type I collagen produces mild
osteogenesis imperfecta.";
J. Biol. Chem. 263:14605-14607(1988).
[21]
SEQUENCE OF 1440-1464 FROM N.A.
MEDLINE=90110490; PubMed=2295701;
Willing M.C., Cohn D.H., Byers P.H.;
"Frameshift mutation near the 3' end of the COL1A1 gene of type I
collagen predicts an elongated pro alpha 1(I) chain and results in
osteogenesis imperfecta type I.";
J. Clin. Invest. 85:282-290(1990).
[22]
SEQUENCE OF 1454-1464 FROM N.A.
PubMed=1995349;
Maatta A., Bornstein P., Penttinen R.P.;
"Highly conserved sequences in the 3'-untranslated region of the
COL1A1 gene bind cell-specific nuclear proteins.";
FEBS Lett. 279:9-13(1991).
[23]
REVIEW ON VARIANTS.
MEDLINE=91184577; PubMed=2010058;
Kuivaniemi H., Tromp G., Prockop D.J.;
"Mutations in collagen genes: causes of rare and some common diseases
in humans.";
FASEB J. 5:2052-2060(1991).
[24]
REVIEW ON VARIANTS.
MEDLINE=97255959; PubMed=9101290;
Kuivaniemi H., Tromp G., Prockop D.J.;
"Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
associated collagen (type IX), and network-forming collagen (type X)
cause a spectrum of diseases of bone, cartilage, and blood vessels.";
Hum. Mutat. 9:300-315(1997).
[25]
- Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
Prockop D.J.;
"Nucleotide sequences of complementary deoxyribonucleic acids for the
pro alpha 1 chain of human type I procollagen. Statistical evaluation
of structures that are conserved during evolution.";
Biochemistry 22:5213-5223(1983).
[7]
SEQUENCE OF 472-607 FROM N.A.
PubMed=2981843;
Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;
"Multixon deletion in an osteogenesis imperfecta variant with
increased type III collagen mRNA.";
J. Biol. Chem. 260:691-694(1985).
[8]
SEQUENCE OF 488-625 FROM N.A.
PubMed=3857621;
Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;
"Intron-mediated recombination may cause a deletion in an alpha 1 type
I collagen chain in a lethal form of osteogenesis imperfecta.";
Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874(1985).
[9]
SEQUENCE OF 1179-1464 FROM N.A., VARIANTS OI-II HTS-1277; ARG-1388 AND
1337-GLU-TYR-1338 DEL, AND VARIANT SER-1434.
MEDLINE=93352646; PubMed=8349697;
Chessler S.D., Wallis G.A., Byers P.H.;
"Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)
chain of type I collagen result in defective chain association and
produce lethal osteogenesis imperfecta.";
J. Biol. Chem. 268:18218-18225(1993).
[10]
SEQUENCE OF 1229-1454 FROM N.A.
TISSUE=bone;
MEDLINE=88124208; PubMed=3340531;
Maekelae J.K., Raassina M., Virta A., Vuorio E.;
"Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
domain.";
Nucleic Acids Res. 16:349-349(1988).
[11]
SEQUENCE OF 1-34 FROM N.A.
MEDLINE=85130970; PubMed=2857713;
Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
"Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
Promoter structure, Alu repeats, and polymorphic transcripts.";
J. Biol. Chem. 260:2315-2320(1985).
[12]
SEQUENCE OF 1-34 FROM N.A.
MEDLINE=88097389; PubMed=3480516;
Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
"Regulatory elements in the first intron contribute to transcriptional
control of the human alpha 1(I) collagen gene.";
Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
[13]
SEQUENCE OF 1-44 FROM N.A.
MEDLINE=88033098; PubMed=2822714;
Rossow C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
de Wet W.;
"DNA sequences in the first intron of the human pro-alpha 1(I)
collagen gene enhance transcription.";
J. Biol. Chem. 262:15151-15157(1987).
[14]
SEQUENCE OF 175-187 AND 274-289.
PubMed=2169412;
Baetge B., Notbohm H., Diebold J., Lehmann H., Bodo M., Deutzmann R.,
Muller P.K.;
"A critical crosslink region in human-bone-derived collagen type I.
Specific cleavage site at residue Leu95.";
Eur. J. Biochem. 192:153-159(1990).
[15]
SEQUENCE OF 263-268.
TISSUE=skin;
MEDLINE=71001508; PubMed=4319110;
Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
"A comparative study of glycopeptides derived from selected vertebrate

REVIEW ON OI VARIANTS.

RX MEDLINE=91374476; PubMed=1895312;

Query Match 59.6%; Score 2765; DB.1; Length 1464;

Best Local Similarity 55.4%; Pred. No. 2.6e-103;

Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGPGTGLPGRGGRGPGSRGFPAGDGVAGPKGAPGGRGPGGAGPKGSGPGEA--- 57
DB 179 GPMGSGRGLPGRGAGPQGFQFPGEPEGPGASGPMGPRGPPGPKNGDDGEAGK 238
QY 58 GRPGAGLP---GAKGLTSGSGSPGP-----DGKTPPGPAGQDGRGPPGPGGAR 105
DB 239 GRPGGRGPPGPGAGRGLGTAGLPMKGRHFGSLDGAKGADGAPGPKGPGSGCENGAP 298
QY 106 GQAGVMGPPGPGAGAGPKAGER-----GVPGPPGAVGPA-----GKDGGAQA 150
DB 299 GQMPRGLPGRGRGAGGAPGARGNDGATGAGPPGTPGAGPPGCAVGAKEAGPQ 358
QY 151 GPPGAPGAGERGEGP-----AGSPGFQGLPGPA 180
DB 359 GPRGSEGGVGEFPGPPGAGAGPAGNPGADGQPGAKGANGAPGAGPFGARGPS 418
QY 181 GPGEAGKEGQGVGDLAGP-----GPSGA-----GEGGPT 213
DB 419 GPGGPGPPGPKNGSEFGAPGSKGDTGAKBPPGVGQPPGAGEGKRGARGEGPT 478
QY 214 GLPGRGGRGSGRGGFGADGAVAGPKGAPGERSGPGAGPKGSGEAGRCEAGLPGAK 273
DB 479 GLPGRGGRGSGRGGFGADGAVAGPKGAPGERSGPGAGPKGSGEAGRCEAGLPGAK 538
QY 274 GLTSGFSGPPGKTKTPGPPAGQDGRPPGPPGARGQAGVMGPPGKAGGEPKAGER 333
DB 539 GLTSGFSGPPGKTKTPGPPAGQDGRPPGPPGARGQAGVMGPPGKAGGEPKAGER 598
QY 334 GVPGPAGVGPAGKDGAGAGQPPGPPAGPAGRGQGPAGSPGFGGLPGAGPPGCEAGKP 393
DB 599 GVPGPAGVGPAGKDGAGAGQPPGPPAGPAGRGQGPAGSPGFGGLPGAGPPGCEAGKP 658
QY 394 GQGVGPDLAGPSPGPAGE-----PGP-----TGLPPOP 423
DB 659 GQGVGPDLAGPSPGARGERFPGGPGVQPPGPPGARGANGAPNDGAKGDAGAP 718
QY 424 GRRGSGRGGFGGADGAVAGPKGAPGERSGPGAGPKGSGE-----AGRCEAGLP 474
DB 719 GSGGAPGLQGMFGERGAALPGKDRDAGPKGADGSGPKDGVRLGTGTPGPPGAP 778
QY 475 GAK-----GLTSGSPSGPDGKTPGPPAGQDGRPPGPPGARGQ----- 515
DB 779 GPKGESGSPGAPGTARGAPGDRGEPGPPGAGFPAGFPAGDQPGKAGGPDGAKGDA 838
QY 516 -----AGVMGFPKPKAGGPKAGRGKAGRGVPPGPA-----VGPAGKDGGAQAQGP 561
DB 839 GPPGAPGAPGPGPIGNVAGKAGSAGPFGATGFFGAAGRVGPPGPGSGNAGPPGP 898
QY 562 GPAGP---AGEGEGQSPGSPGQGLPGPAGPPEAGKFGEGVPGDLGAPGSPGAPGE 618
DB 899 GPAGKEGKGPRGRTGAPRGPEVPPGPPGAGEKSGFGADGAPAGPTGPPQGIQOR 958
QY 619 GTGLPGRGGRGPGSRGFPAGDGVAGPKGAPGARGSGPPAGP---KGSCEAGRPGEA 675
DB 959 GVVGLPGQGERGFPGLPGSPGPKQGPSGASGERGPPGMPGPPGLAGPPGSGREGAP 1018
QY 676 GLPGAKGLTSGSPGPDGKTPGPPAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEP 735
DB 1019 GAEPSGRDGSFGAKGDRGETGAPGPPGAPGAPGPPVGPAGKSGDRGETGPPAGAPV 1078
QY 736 GKAGERGVPPGPPGAVGAPGKDG-----AGAQGPFGP-----AGFA 771
DB 1079 GPAGAGGAPGPPGPRDKGKETGEQDGRGKGRHGFSGLQPPGPPGSPGEGQPSGASGFA 1138
QY 772 GERGEGCPAGSP---GFQGLPGPAGPPGAGKFGGQGVPGDLGAPGSPGAG 820

Db 1139 GPRPPGCSAGAPGKGLNGLPPIGPPGRGRTGDAGPVGPPGPPGPPG 1190
RESULT 2
O76045
ID O76045 PRELIMINARY; PRT; 1461 AA.
AC O76045;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Pro alpha 1(I) collagen.
GN Names:COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88329734; PubMed=2843432;
RA D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene.";
RL Gene 67:105-115(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91138770; PubMed=1995349;
RA Maatta A., Bornstein P., Penttinen R.P.;
RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";
RL FEBS Lett. 279:9-13(1991).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=92157916; PubMed=1787829;
RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL Matrix 11:375-379(1991).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=98107942; PubMed=9443882;
RA Korhko J., Ala-Kokko L., De Paep A., Nuytinc L., Barley J.,
RA Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
RT scanning by conformation-sensitive gel electrophoresis identifies only
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I;
RT identification of common sequences of null-allele mutations.";
RL Am. J. Hum. Genet. 62:98-110(1998).
RN [7]
RP SEQUENCE FROM N.A.
RA Korhko J.M., Barley J.J., Nuytinc L., DePaep A., Prockop D.J.,
RA Ala-Kokko L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017178; AB94054.2;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.


```

DR InterPro: IPR008161; Clg helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 18.
DR ProDom: PD000007; Clg helix; 3.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR SMART: SM00214; VWC; 1.
DR PROSITE: PS01208; WFC1; 1.
DR PROSITE: PS0184; WFC2; 1.
KW Collagen.
SQ SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30EA78E21 CRC64;

Query Match          59.5%; Score 2761; DB 2; Length 1461;
Best Local Similarity 55.3%; Pred. No. 3.8e-103;
Matches 560; Conservative 31; Mismatches 229; Indels 192; Gaps 19;

QY 1 GPPGPGTGLPGRGPGGSRGPPGADGVAGKPGAGRGSGPAGPKGSGPGEA--- 57
DB 176 GPMGSGRGLGPPGARGPQGFQGPGEPPGPGASGPMGPRGPPGPKNGDDGEAGKP 235
QY 58 GRPGAGLP---GAKLTGSPSGP-----DGKTPPGPAGDGRGPPGPPGARG 105
DB 236 GRPGERGPPGQARGLPGTAGLPGMKHGRGSGLDGAKDAGPAGPKGEFGSGENGAP 295
QY 106 GQAGVMGPPGPKGAAGEPKAGER-----GVPGPPGAVGPA-----GKDGEGAGQ 150
DB 296 GQMGPRGLGEGRGPGAPGAPGARGNDGATGAAGPPGTGTPAGPPGFCVAGKAGEAGPQ 355
QY 151 GPPGAPGAGERGEGCP-----AGSGPFGQLPGPA 180
DB 356 GPRGSEGGQVGRGEPGPPGPAAGPAGNPGADGQPGKANGAGPAGIAGAPGFGARGPS 415
QY 181 GPPGAGKGEQGVPGDLGAP-----GSGGPA-----GEGCPT 213
DB 416 GQGGPGPPGPKNGSGEPGAPGSKDGTAKGEPGVGVQPPGAGEGKRGAGEGPT 475
QY 214 GLPFPGERGSGRFGPGADGVAGPKGFPAGRGSGPAGPKGSGPGEAGRPGEGAGLPCAK 273
DB 476 GLPFPGERGSGRFGPGADGVAGPKGFPAGRGSGPAGPKGSGPGEAGRPGEGAGLPCAK 535
QY 274 GLTSGSGSPGPKTTPCPACDGRPPPPGARGQAGVMPGPKGAGEGPKAGER 333
DB 536 GLTSGSGSPGPKTTPCPACDGRPPPPGARGQAGVMPGPKGAGEGPKAGER 595
QY 334 GVPGPPGAVGPKDGEAGAGPAGPAGRGSGPAGSGPFGQLGPPAGPGEAGKP 393
DB 596 GVPGPPGAVGPKDGEAGAGPAGPAGRGSGPAGSGPFGQLGPPAGPGEAGKP 655
QY 394 GQGVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 423
DB 656 GQGVPGDLGAPGSGAGERGFPGERGVQPPGPRGANGAPGNDGAKGDAGAGPAP 715
QY 424 GERGGGSRGPPGADGVAGPKGFPAGRGSGPAGPKGSGPGE-----AGRPGAGLP 474
DB 716 GSQAFGLQGMPEERGAAGLPKPKGDRDAGPKGADGSGGKDGVLGTGTP:GPPGAPAP 775
QY 475 GAX-----GLTSGSGSPGPKTTPCPACDGRPPPPGARGQ----- 515
DB 776 GDXSGSGSPGAPGARGAPGDRGEPGPPGAPGAGPDAGDQGFAGKGEFGDAGAGDA 835
QY 516 -----AGVMGPPGKGAAGEGPKAGERGVPPGPA-----VGPAGKDEAGAGP 561
DB 836 GPPGAPGAPPEPILNVAPKAGKAGSAPPATGPPGAAGRVGPPGSGNAGPPGPP 895
QY 562 GPAGP---AGERGEGPAGSPFQGLPAGPAGGPKGEGQGVGDLGAPGSGPAGP 618
DB 896 GPAGKEGKGKPRGTGTGPRGVEVPPGPPGAGEKGSFGADGAPAGTGTGPGQIAGQR 955
QY 619 GPTGLPGRPGRGSGRFGFCADGVAGPKPAGERGSPGAPG---KGSFGEAGRGEA 675

```

Db 956 GVVLPGQGRGFRGLPGSPGSPGKQSGASGERGPPGMPGPPCLAGPAGSGREGAP 1015

Qy 676 GLFGAKGLTGSPGSPGPDGKTPGPPGAGQDGRPPGPPGARGQAGVMPGPPKGAAGEP 735

Db 1016 GAEGSPRGDSGPKAGKDRGTGTPAGPPGAPGAPGAPGVPAGKSGDRGTGTPAGPAPV 1075

Qy 736 GKAGERGVPPGCAVGPAGKDGSE-----AGAOGPPGP-----AGPA 771

Db 1076 GPVGARGPAGPQGRDGTGEGQDRGIXHGRGSLGQFPGPPSPGQSPGSGASGPA 1135

Qy 772 GERGEGPAGSP---GFGGLPGPAGPAGKPGEGQGVGFDLGAOPSPGAP 820

Db 1136 GPRGPPGSGAGAPKDGGLGPGIPGPPGPRGTGDAGVPVGPVPPGPPGPPG 1187

RESULT 3

Q8N473 PRELIMINARY; PRT; 1464 AA.

AC Q8N473;

DT 01-OCT-2002 (T-EMBLrel. 22, Created)

DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)

DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)

DE Alpha 1 type I collagen, preproprotein.

GN Name=COL1A1;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

FN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Scapleton M., Soares M.B., Donald M.P., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellanc N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Murny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E., Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Strausberg R.;

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR ENBL; BC035331; AAH3533.1; --

DR GO; GO:0005581; C:cytoplasm; IEA.

DR GO; GO:0005737; C:cytoplasm; IEA.

DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.

DR GO; GO:0006817; P:phosphate transport; IEA.

DR InterPro: IPR008161; Clg helix.

DR InterPro: IPR008160; Collagen.

DR InterPro: IPR000885; Fib_collagen_C.

DR InterPro: IPR009041; PMP_SGCI.

DR InterPro: IPR001007; VWF_C.

DR Pfam; PF01410; COLFI; 1.

DR Pfam; PF01391; Collagen; 18.

DR ProDom; PD000007; Clg helix; 3.

DR ProDom; PD002078; Fib_collagen_C; 1.

DR SMART; SM00038; COLFI; 1.

DR SMART; SM00214; VWC; 1.

FT MOD_RES 1160 1160 3-hydroxyproline (By similarity).
 FT CARBOHYD 261 261 O-linked (Gal...) (By similarity).
 FT CARBOHYD 1361 1361 N-linked (GlcNAc...) (By similarity).
 FT VARIANT 208 208 G -> A (in OI; severe).
 SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;

Query Match 59.5%; Score 2759; DB 1; Length 1460;
 Best Local Similarity 55.9%; Pred. No. 4.5e-103;
 Matches 559; Conservative 31; Mismatches 230; Indels 180; Gaps 19;

QY 1 GPPGPGPTGLPQPCERGGPGS-----RGPPGA-----DGVAGPKPAGERGSGPGA 48
 Db 187 GPPGAGPQGFQPCGCEPCASGPMGRPPGPPGKNGDDGEAGKPGRGGERGPPGPQ 246
 QY 49 GPKGPGEGAGRP-----GEAGLPKAGKLTGSPGSGPPGKTTGPPGAGQD 93
 Db 247 GARGLFTAGLPQMKHGRFSLDGAAGKADAGPAGPKBPGSPGNGAPGQMGPRGLPGER 306
 QY 94 GRGPGPPGARGQ---AGVMGPPGPKGAAGB---PGKAGRGVPPGPAVGAPGAKGCEA 147
 Db 307 GRPGAPGACAKNDGATGAACPPPTGTPAGPPFGAVGAKGEAGPGARGSGSPQGV 366
 QY 148 GAQGGPPGAPGAGERGEOGP-----AGSPGFQGLPFPAGPPGEGEAGKPGEQ 192
 Db 367 GEPGPPGAGAGPAGNPGADQCPGAKGANGAPGIAGAPFGPARGPSPGQSPGPPGPK 426
 QY 193 VYPGDLGAP-----GPSGA-----GEPGTLGPPGGERGGP 225
 Db 427 GNSGPGAPGNKGDGAKGEPGPTGIQPPGAGBEGKRGARGEPGTLGPPGGERGGP 486
 QY 226 GSRGPPGADVAGPKPAGERGSPGAPGKSPGEGAPGEGAGLPGAKGLTGSFGSPGD 285
 Db 487 GSRGPPGADVAGPKPAGERGSPGAPGKSPGEGAPGEGAGLPGAKGLTGSFGSPGD 546
 QY 286 GKTGPPGAGQDRGPPGPPGARGQAGVMGPPGPKGANGEPGKAGRGVPPGPAVGPA 345
 Db 547 GKTGPPGAGQDRGPPGPPGARGQAGVMGPPGPKGANGEPGKAGRGVPPGPAVGPA 606
 QY 346 GKDAGAGAGPPGAPGAGERGEOGPAGSPGQGLPFPAGPPGAGKPGEGVPGDLGAP 405
 Db 607 GKDAGAGAGPPGAPGAGERGEOGPAGSPGQGLPFPAGPPGAGKPGEGVPGDLGAP 666
 QY 406 GFSGPAGE-----PGP-----TGLPDPGPPGEGGSGRGFP 435
 Db 667 GFSGARGGERGPPGGERGVGPPGPPGAPGANGAPNGDAGKADAGAPGAGPQAGPAGQMP 726
 QY 436 GADGVAGPKPAGERGSPGAPGKSPGEGAGLPGAKGLTGSFGSPGD 477
 Db 727 GERGAGLPGPKGDRGDAGPKGADGSPKDGVRGLTGTGIPGPPGAPAGPDDTGEAGSPGA 786
 QY 478 GLTSPGSPGDKTGTGPPGAGQDRGPPGPPGARGQ-----AGVMGFP 522
 Db 787 GPTGARGAPDRGEPGPPGAPGAGFPAGADGQPKAGBPGDAGAKGDAGPPGAGTGP 846
 QY 523 GPKGAAGBPKAGRGVCPGCA-----VGPAGKDGAGAGQPPGAPG---AGER 570
 Db 847 GPTGNVAGPPKARGSGAPGATGFPAGAGRVGPPGSGNAGPPGPPGAKGEGGARG 906
 QY 571 GQGPAGSPGFLPGLPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 630
 Db 907 GETGAPRGVGVGPPGPPGAGEKSPGADGAPAGPAGTGGIAGQGVVGLFGQGER 966
 QY 631 GPGSGRPPGADVAGPKPAGERGSPGAPG---KSPGEGAPGEGAGLPGAKGLTGP 687
 Db 967 GPPGLPGSGEPGKSPGSGTSGERGPPGPMGPPGLAGPPGSGREGSPGAGSGPDRGSP 1026
 QY 688 GSPGDKTGTGPPGAGQDRGPPGPPGARGQAGVMGPPGPKGANGEPGKAGRGVPPGP 747
 Db 1027 GPKGDRGETGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 1086
 QY 748 GAVGAPGAKDGE-----AGAQPDPG-----AGPAGERGEGGAPGSP 783
 Db 1087 GPRGDKGTGEOGDRGKIHGRFSLQPPGPPGSGPQSPGSGASGAPGPRGPPGSGSP 1146

QY 784 ---GFGGLPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 820
 Db 1147 GKDLNGLPPIGPPGPRGRTGDAGVPGPFPFPFPFPFPFPFPFPFPFPFPFPFP 1186

RESULT 5

Q63079 PRELIMINARY; PRT; 1453 AA.
 AC Q63079;
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)
 DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
 DE Collagen alpha1 (Fragment).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Bone, and Tooth;
 RX MEDLINE=99163824; PubMed=10065941;
 RA Brandsten C., Lundmark C., Christersson C., Hammarstrom L., Wurtz T.;
 RT "Expression of collagen alpha1(I) mRNA variants during tooth and bone
 formation in the rat";
 RL J. Dent. Res. 78:11-19(1999).
 DR EMBL: Z78279; CAB01633.1; -;
 DR GO: GO:0005581; C:collagen; IEA.
 DR GO: GO:0005737; C:cytoplasm; IEA.
 DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR GO: GO:0006517; P:phosphate transport; IEA.
 DR InterPro: IPR008161; Clg helix.
 DR InterPro: IPR008150; Collagen.
 DR InterPro: IPR000885; Fib collagen_C.
 DR InterPro: IPR009041; PMP_SGCI.
 DR InterPro: IPR001007; WFC_C.
 DR Pfam: PF01410; COLFI; 1.
 DR Pfam: PF01391; Collagen; 18.
 DR ProDom: PD000007; Clg helix; 3.
 DR ProDom: PD002078; Fib collagen_C; 1.
 DR SMART: SM00038; COLFI; 1.
 DR SMART: SM00214; WVC; 1.
 DR PROSITE: PS01208; WFC_1; 1.
 DR PROSITE: PS0184; WFC_2; 1.
 KW Collagen.
 FT NON TER.
 SQ SEQUENCE 1453 AA; 137886 MW; E6896BDC19A4A1D8 CRC64;

Query Match 59.2%; Score 2748.5; DB 2; Length 1453;
 Best Local Similarity 52.3%; Pred. No. 1.2e-102;
 Matches 556; Conservative 33; Mismatches 231; Indels 243; Gaps 19;

QY 1 GPPGE---PGPTGLPQPCERGGPSRGFP-----ADGVAGPKPAGER 42
 Db 117 GPPGQGIISGQGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGLPGL 176
 QY 43 GSPGAPGPKSGPGEAGRPGEAGLPGAKGLTGSFGSPGDKTGTGPPGAGQDRP---GPP 99
 Db 177 GLPFPGAPGPPGQFPQFPQFPQFPQFPQFPQFPQFPQFPQFPQFPQFPQFPQFPQ 236
 QY 100 GPPGARGQAGVMGFP-----GPKGAGEGPK-----A 126
 Db 237 GPGAGLPGTGLPQMKHGRFSLDGAAGKGTGTPAGPKGEGSGENGTPQMGPRGLP 296
 QY 127 GERGVPPGAVGPAKKGCEAGAGQPPG-----AGPAGERGEOGP- 167
 Db 297 GERGPPGPGTAGAGNDGAVGAAGPFGTGTGPPFPFGAAGKAGEAGPQARGSEGP 356
 QY 168 -----AGSPGFGGLPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 189
 Db 357 GVRGPPGPPGAPAGPAGNPGADGQPKAGKANGAPGIAGAPGFGKAGPSGPGPSPGAP 416
 QY 190 GEQGVGDLGAP-----GPSGA-----GEPGTLGPPGGER 222

```
Db 417 GPKGTSGPAGFNKNGTGAKEGPGAGVQGPAGBEGRKRGARGEPGSLPGPPGER 476
Qy 223 GPGSRGPGADGAGVAGPKPAGERGSGPCAPGKSPGCEAGPPEAGLPGAKGLTGSQSP 282
Db 477 GPGSRGPGADGAGVAGPKPAGERGSGPCAPGKSPGCEAGPPEAGLPGAKGLTGSQSP 536
Qy 283 GPDGKTGPPGAGDGRGPPGPPCARCGAGVMGPPGPKAAGPFGKAGRGVGPFGAV 342
Db 537 GPDGKTGPPGAGDGRGPPGPPCARCGAGVMGPPGPKAAGPFGKAGRGVGPFGAV 596
Qy 343 GPAGKDGAGAGQPPGPPGAPGAGRGEGQPGAPSPFOGLPGPAGPBGAGKPGEGVPGDL 402
Db 597 GPAGKDGAGAGQPPGPPGAPGAGRGEGQPGAPSPFOGLPGPAGPBGAGKPGEGVPGDL 656
Qy 403 GAPGSGPAGE-----PGP-----TGLPDPGPPGGRGGPSR 432
Db 657 GAPGSGGARGEGPPGEGVQGPAGPGRNGNAGPNDGAKGTGAPGAPSGAGPLQ 716
Qy 433 GPPGADGAGVAGPKPAGERGSGPPGAPKSPGE-----AGPBGAGLPGAK----- 477
Db 717 GMPGEGAGAGLPGPKGDRDAGPKGADGSPKDGVRGLTGIPGPPGAPGAGDGEAGPS 776
Qy 478 ---GLTSGSPGPPGDKTGPAGQDGRGPPGPPGARGO-----AGVM 519
Db 777 GPAGTGTARGAPGDRGEGAGPPGAPGAGFAGPAGDQGPAGKGPFGTGVKGDAGPPGAPGA 836
Qy 520 GPPGPKGAAGPFGKAGRGVFPFGA-----VGPAGKDGAGAGQPPGPPGAP--A 567
Db 837 GPPGPIGNVAGPFGKSGRGAAGPGATGFPGAAGRVGPPGSGNAGPPGPPGVKSGGK 896
Qy 568 GERGQGAGSPGFOGLPGPAGPPEAGKPGEGVPGDLGAPGSGPAGEPGPTGLPGPP 627
Db 897 GPRGTGTAGRPGEVGPFGPPGAPGEXGSPADGPAGSPGPTGPGQIAGQGVVGLPGQR 956
Qy 628 GERGPGSRGPPGADGAGPKGAPGERSGPPGAP---KSGSPGAGRPGEAGLPGAKGLT 684
Db 957 GKRGPGLPGSGEGPKQPSGASGERGPPGPMGPPGLAGPPGSGREGSGAGSPGRD 1016
Qy 685 GSPGSPGDPKGTGPPGAGQDGRGPPGPPGARGQAGVMGPPGKGAAGEPKAGRGV 744
Db 1017 GAPGAKGRGRTGTPAGPPGAPGAPGAPGVPVAGKNGDRGTGTPAGPAGTGPAGARGPA 1076
Qy 745 GP-----PGAVPAGKDGAGAGQPPGPPGAPGAGRGEGQPA 780
Db 1077 GPGPRGDKGTGEGDGRGKRGHGFSLQPPGSPGPGCGGSGASGAPGPPGSA 1136
Qy 781 GSP---GFGPLGPPAGPPEAGKPGEGVPGDLGAPGSPGAP 820
Db 1137 GSPGKDGGLNGLPGIPGPPGPRGRTGDSGPPGPPGPPGPPG 1179

RESULT 6
CALL MOUSE
ID CALL MOUSE STANDARD; PRT; 1453 AA.
AC P11087; Q60635;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=Colla1; Synonyms=Colal;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EVB/N;
RA MEDLINE=96033240; PubMed=8535610;
RX Li S.W., Khillan J., Prockop D.J.;
RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
of type I procollagen.";
RL Matrix Biol. 14:593-595 (1995).
```

```
RN [2]
RP SEQUENCE OF 518-1128 FROM N.A.
RX MEDLINE=86137403; PubMed=3841523;
RA French B.T., Lee W.-H., Maul G.G.;
RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I) collagen
protein.";
RL Gene 39:311-312 (1985).
RN [3]
RP SEQUENCE OF 735-1130 FROM N.A.
RX MEDLINE=83141374; PubMed=6298597;
RA Monson J.M., Friedman J., McCarthy B.J.;
RT "DNA sequence analysis of a mouse pro alpha 1(I) procollagen gene:
evidence for a mouse B1 element within the gene.";
RL Mol. Cell. Biol. 2:1362-1371 (1982).
RN [4]
RP SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX MEDLINE=83157109; PubMed=6219867;
RA Monson J.M., McCarthy B.J.;
RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
evidence for insertions or deletions in gene coding sequences.";
RL DNA 1:59-69 (1981).
RN [5]
RP SEQUENCE OF 1442-1453 FROM N.A.
RX MEDLINE=88124276; PubMed=3340560;
RA Mooslehner K., Harbers K.;
RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size of
the 3'-untranslated region.";
RL Nucleic Acids Res. 16:773-773 (1988).
CC -1- FUNCTION: Type I collagen is a member of group I collagen
(fibrillar forming collagen).
CC -1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
bones. In bones the fibrils are mineralized with calcium
hydroxyapatite.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- SIMILARITY: Belongs to the fibrillar collagen family.
CC -1- SIMILARITY: Contains 1 VWFC domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

DR PROSITE; PS01208; WFC1; 1.
 DR PROSITE; PS0184; WFC2; 1.
 KW Collagen; Connective tissue; Extracellular matrix; Glycoprotein;
 KW Hydroxylation; Pyroglutamate; Carboxylic acid; Repeat; Signal.
 FT SIGNAL 1 22 N-terminal propeptide
 FT PROPEP 23 151 Collagen alpha 1(I) chain.
 FT CHAIN 152 1207 C-terminal propeptide.
 FT PROPEP 1208 1453 WFC.
 FT DOMAIN 29 87 Nonhelical region (N-terminal).
 FT DOMAIN 152 167 Triple-helical region.
 FT DOMAIN 168 1181 Nonhelical region (C-terminal).
 FT DOMAIN 1182 1207 Pyroglutamate carboxylic acid (By
 FT MOD_RES 152 152 similarity).
 FT MOD_RES 160 160 Allylsine (By similarity).
 FT MOD_RES 254 254 5-hydroxylysine (By similarity).
 FT MOD_RES 1153 1153 3-hydroxyproline (By similarity).
 FT CARBOHYD 56 56 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 254 254 O-linked (Gal...) (By similarity).
 FT CARBOHYD 1354 1354 N-linked (GlcNAc...) (By similarity).
 FT SITE 734 736 Cell attachment site (Potential).
 FT SITE 1082 1084 Cell attachment site (Potential).
 FT CONFLICT 1450 1450 A -> V (in Ref. 5).
 SQ SEQUENCE 1453 AA; 137944 MW; 3B802E535DF81808 CRC64;

Query Match 58.9%; Score 2732.5; DB 1; Length 1453;
 Best Local Similarity 54.7%; Pred. No. 5:2e-102;
 Matches 547; Conservative 33; Mismatches 240; Indels 180; Gaps 16;

QY 1 GPPGPGPTGLPGPGE-----RCGPGSRGPGADGVAGPKPAGERGSGPAA 48
 DB 180 GPPGPGPGCGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPG 239
 QY 49 GPKGPGGAGRP-----CEAGLPGAKGLTSGSPGPGDGKTPGPGAGQD 93
 DB 240 GARGLPFGTAGLPGMKHGRFSGLDGAKGDAGPAGKPGFSGNGAPGQNGPLGGER 299
 QY 94 GPGPGPGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 147
 DB 300 GPGPGPGTAGGNDGAVGAGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPG 359
 QY 148 GAGPGPGPAGPAGERGEGP-----AGSPGPGGLPGPAGPGEAGKPGGEQ 192
 DB 360 GEPGPGPAGAGPAGNPGCADQPGKAGANGAPGAGPAGPAGPAGPAGPAGP 419
 QY 193 GVPGLDGP-----GPSGPA-----GPGPTGLPGPGERGGP 225
 DB 420 GNSGPGAPGKNGDTGAKGPGATGQVQPPGPGAGEGKRGAGPFGSGLPFGPGERGP 479
 QY 226 GSRGPGADGVAGPKPAGERGSPGAPGKSPGEPAGEPGEAGLPAGKGLTSGSPGPD 285
 DB 480 GSRGPGADGVAGPKPAGERGSPGAPGKSPGEPAGEPGEAGLPAGKGLTSGSPGPD 539
 QY 286 GKTGPGGAGDGRPGPGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 345
 DB 540 GKTGPGGAGDGRPGPGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 599
 QY 346 GKDGEAGAGQPPGPGPAGERGEGQPGSPGQGLPGPAGPAGPAGPAGPAGP 405
 DB 600 GKDGEAGAGQPPGPGPAGERGEGQPGSPGQGLPGPAGPAGPAGPAGPAGP 659
 QY 406 GPSGAGE-----PGP-----TGLPGPGERGPGSGPFP 435
 DB 660 GPSGARGRGPGERGVGPGPGPAGPRNNGAPNDGAKGDTGAPGPGSGAPGLOGMP 719
 QY 436 GADGVAGPKPAGERGSPGAPGKSPGEPAGEPGEAGLPAGKGLTSGSPGSP 486
 DB 720 GERGAAGLPKPGKDRGDAGPKGADGSPKDGARGLTGIPGPGPAGPAGKGEAGPSP 779
 QY 487 GPD-----GKTGPGPAGDGRPGPAGPAGPAGPAGPAGPAGPAGPAGP 522
 DB 780 GPTGARGAPGDRGEAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 839

QY 523 GPKGAAGEPKKAGERGVGPPGGA-----VGPAGKDGAGAGQPPGPGPAGP---ACGR 570
 DB 840 GPTGNVGAEPKPRGAGPGATGPGGAAGVGPVPSGNAGPFPGPVKGKGR 899
 QY 571 GEOGPAGSPGQGLPGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 630
 DB 900 GETGPAGRGEVGPFPFPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 959
 QY 631 GPGSGSGFPAGDGVAGPKPAGERGSPGAGP---KSGSGEAGRGEAGLPAGKGLTSP 687
 DB 960 GFGPLPGSGEPKQKPGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 1019
 QY 688 GSGPGDKTKGPGPAGDGRGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 746
 DB 1020 GAKGDRGETGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1079
 QY 747 -----PGAVGPAGKDGAGAGQPPGPGPAGPAGPAGPAGPAGPAGPAG 783
 DB 1080 GPRGDKGETGEQDGRGKGRGFGSLGQPPGSGSGSGSGSGSGSGSGSGSG 1139
 QY 784 ---GFGGLPGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 820
 DB 1140 GKDLGLNGLFPGP-GPPGPRGTGDSGPAGPAGPAGPAGPAGPAGPAGP 1179

RESULT 7
 Q810J9 PRELIMINARY; PRT; 1453 AA.
 ID Q810J9
 AC Q810J9;
 DT 01-JUN-2003 (T-EMBLrel. 24, Created)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
 DT 01-VAR-2004 (T-EMBLrel. 26, Last annotation update)
 DE Procollagen, type I, alpha 1.
 GN Names=Collal;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=FVB/N; TISSUE=Colon;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Fellin F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Klausner R.D., Collins F.S., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Altschul S.F., Zeeberg B., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udgin T.B., Roschlyki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC050014; RAH50014.1; -;
 DR MGI; MGI:89467; Collal
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005737; C:cyclopia; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.

```
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib collagen_C.
DR InterPro; IPR009041; BMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS01184; VWF_C_2; 1.
DR Collagen.
KW SEQUENCE 1453 AA; 138032 MW; 0B7F06BBB9A1D5EA CRC64;

Query Match 58.9%; Score 2732.5; DB 2; Length 1453;
Best Local Similarity 54.7%; Pred. No. 5.2e-102;
Matches 547; Conservative 33; Mismatches 240; Indels 180; Gaps 16;

QY 1 GPPGPGTGLPDPGE-----RGPGSRGPPGADGVAGPKGAPAGERGSGPA 48
DB 180 GPPGAPGPGQFGPGGPGGPGGPMGPRGPPGPKNGKNGDDGAGKPRGPRGPRGPPQ 239
QY 49 GPKSGPGEAGRP-----CEAGLPKAGLTSGPSGSPGDPKGTGPPGPAQD 93
DB 240 GARGLPGTAGLPKMKHGRGFGSLDGAAGDAGAPKGPSPGFGNGAPGQMGPLGPR 299
QY 94 GRPGPPGPGARGQAGVGMFGPPKGAAGE-----PGKAGRGVPGPPGAVGPAKGDGEA 147
DB 300 GRPGPPGTAGARGNDGAVGAAGPPGTGTPGPPFGAVGAKGAGPQARGSGPQVR 359
QY 148 GAGGPPGAPAGERGEGP-----AGSPGQGLPGPAGPPGEGAGKPGEQ 192
DB 360 GEPGPPGAPAGPAGNPGADQPGKAGANGAPFPFGARGSPGPGSGPPGPK 419
QY 193 GVPGLDLAGP-----GPSGA-----GRPGTGLPDPGEGGP 225
DB 420 GNSGEPGAPGNKDTGAKGEGCATGVQPPFAGEEGKRGAGPFGSLGPPGERGP 479
QY 226 GSRGPPGADGVAGPKPAGERGSPGAPGKSPGEGAPGPEAGLPKAGLTSGPSGPD 285
DB 480 GSRGPPGADGVAGPKPAGERGSPGAPGAPGAPGKSPGEGAPGPEAGLPKAGLTSGPSGPD 539
QY 286 KHTGPPGAGDGRGPPGPPGARGOAGVMGPPGKGAAGSPGKAGRGVGPAGVGA 345
DB 540 KHTGPPGAGDGRGPPGPPGARGOAGVMGPPGKGTAGTGFKAGRGVGPAGVGA 599
QY 346 GKDGEAGAGPPGAPAGERGEGQPGAPSPGQGLPGPAGPPGEGAGKFGEGQVFDLGA 405
DB 600 GKDGEAGAGAGPPGAPAGERGEGQPGAPSPGQGLPGPAGPPGEGAGKFGEGQVFDLGA 659
QY 406 GPSGAGE-----PGP-----TGLPGERGGRGSGRP 435
DB 660 GPSGARGERGPPGERGVQPPGPPGFRNGNAPGNDGAKGDTGAPGAPGSGAPGLQMP 719
QY 436 GADGVAGPKPAGERGSPGAPGKSPG-----AGRPGEAGLPKAGLTSGPS 486
DB 720 GEGGAGLPKPGKDRGDAGPKADSGPKDARGLTGTPGPPGAGAGDKGEGAPSGPP 779
QY 487 GPD-----GKTGPPGAGDGRGPPGPPGARGQ-----AGVMGFP 522
DB 780 GPTGARGAPDRGEAGPPGAPGAPGADGQFGAKGPFDDTVKNGDAGPPGAPGAP 839
QY 523 GPKGAAGEPKAGRGVPPGCA-----VGPAGKDGAGAGGPPGAPG---AGER 570
DB 840 GPIGNVAGPKPGRGAAGPPGATGPPGAAGRVGPPGSGNAGPPGPPGVKGGKGR 899
QY 571 GEQGAGSPGFGGLPFPAGPPGEGAGKPGQVPGDLGAPGSPGAPGEPCTGLPQPGR 630
DB 900 GETGPAGRGVEGPPGPPGAPGKSPGADGAPGAPGTPGPGIAGQGVVGLPQORGER 959

RESULT 8
Q6LAN8 PRELIMINARY; PRT; 1069 AA.
AC Q6LAN8
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TEMBLrel. 27, Last annotation update)
DE Collagen type I alpha 1 (Fragment).
GN NamesCOL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97141927; PubMed=8988177;
RA Simon M., Pedoutour F., O'Brien K., Sirvent N., Kedra D., Guilbaud C.,
RA Turc-Carel C., Dumanski J.P.;
RT "Regulation of the platelet-derived growth factor B-chain gene via
RT fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans
RT and giant-cell fibroblastoma.";
RL Nat. Genet. 15:95-98 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RA O'Brien K.P.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
EMBL; X98705; CAA67261.1; -.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR009041; BMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01391; Collagen; 15.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg_helix; 3.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS01184; VWF_C_2; 1.
DR Collagen.
FT NON TER 1069 1069
SQ SEQUENCE 1069 AA; 97445 MW; EE279B10572FB980 CRC64;

Query Match 57.8%; Score 2681.5; DB 2; Length 1069;
Best Local Similarity 58.4%; Pred. No. 4.6e-100;
Matches 524; Conservative 28; Mismatches 229; Indels 117; Gaps 10;

QY 1 GPPGEGTGLPFPGERGPPGSRGPPGADGVAGPKPAGERGSPGAPGKSPGAGRP 60
DB 203 GPPGEGPGEASGPMGPRGPPGPPGKNGDDGEGAGKFGFRGPPGPPGQARGLPGTAGLP 262
QY 61 GEAGLPKAGLTSGPSGPPGDKTGTGPPGAGDGRPPGPPGARGOAGVMGPPGPKGAA 120
DB 263 GMKGRGFGSLDGAAGDAGAPGKGFSGPENGAPGQMGPRGLPGERGPPGAPGAGAR 322
QY 121 GEPKAGRGVPPGPAVGP-----AGKDGEAGAGPPGPPGAPGAGERGQGP----- 167
DB 323 GNDGATGAAGPPGTGTGPPGPPGPPGAVGAKGAGPQGPGRSGEGPQGVGPPGPPGAGAA 382
```


DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Gallus gallus (Chicken).
OC Sukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]_TaxID=9031;
RP SEQUENCE OF 1-153 FROM N.A.
RX MEDLINE=88056316; PubMed=3678834;
RA Finer M.H., Boedtker H., Doty P.;
RT "Construction and characterization of cDNA clones encoding the 5' end
of the chicken pro alpha 1(I) collagen mRNA.";
RL Gene 56:71-78(1987).
RN [2]
RP SEQUENCE OF 1-144 FROM N.A.
RX MEDLINE=88007542; PubMed=2820966;
RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RT "Unusual DNA sequences located within the promoter region and the
first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [3]
RP SEQUENCE OF 152-1187.
RX MEDLINE=82231995; PubMed=7093229;
RA Higberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M., Kang A.H.,
Gross J.;
RT "Amino acid sequence of chick skin collagen alpha 1(I)-C88 and the
complete primary structure of the helical portion of the chick skin
collagen alpha 1(I) chain.";
RL Biochemistry 21:2048-2055(1982).
RN [4]
RP SEQUENCE OF 1200-1205.
RX MEDLINE=72243016; PubMed=5047697;
RA Eyre D.R., Glimcher M.J.;
RT "Evidence for a previously undetected sequence at the carboxyterminus
of the alpha 1 chain of chicken bone collagen.";
RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN [5]
RP SEQUENCE OF 981-1453 FROM N.A.
RX MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
including the carboxy-terminal propeptide sequences.";
RL Biochemistry 20:996-1006(1981).
RN [6]
RP SEQUENCE OF 1311-1453 FROM N.A.
RX MEDLINE=80134546; PubMed=6987088;
RA Showalter A.M., Resciotta D.M., Eikenberry R.F., Yamamoto T.,
Pastan I., Decrombrughe B., Fietzek P.P., Olsen B.R.;
RT "Nucleotide sequence of a collagen cDNA-fragment coding for the
carboxyl end of pro alpha 1(I)-chains.";
RL FEBS Lett. 111:61-65(1980).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
(fibrillar forming collagen).
CC -!- SUBUNIT: Trimer of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
bones. In bones the fibrils are mineralized with calcium
hydroxyapatite.
CC -!- PTM: Proline residues at the third position of the tripeptide
repeating unit (G-X-Y) are hydroxylated in some or all of the
chains. Pro-1153 is the only 3-hydroxypro and the only
hydroxylated proline in position X.
CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
CC -!- SIMILARITY: Contains 1 WFCC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; M17839; AAA48704.1; -.
DR EMBL; M17838; AAA48704.1; JOINED.
DR EMBL; V00401; CAA23695.1; -.
DR EMBL; M10571; AAA48671.1; ALT_SEQ.
DR EMBL; M17607; AAA48672.1; -.
DR PIR; A27179; A27179.
DR PIR; I50629; I50629.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib_collagen_C.
DR InterPro; IPR01007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS0184; VWF_C; 2; 1.
KW Collagen; Connective tissue; Direct protein sequencing;
KW Extracellular matrix; Glycoprotein; Hydroxylation;
KW Pyrrolidone carboxylic acid; Repeat; Signal.
FT SIGNAL 1 22
FT PROPEP 23 151 N-terminal propeptide.
FT CHAIN 152 1205 Collagen alpha 1(I) chain.
FT PROPEP 1206 1453 C-terminal propeptide.
FT DOMAIN 31 89 VWF_C
FT MOD_RES 152 152 Pyrrolidone carboxylic acid.
FT MOD_RES 160 160 Allysine (By similarity).
FT MOD_RES 254 254 5-hydroxylysine (By similarity).
FT MOD_RES 851 851 5-hydroxylysine (Potential).
FT MOD_RES 1081 1081 Hydroxyproline (Potential).
FT MOD_RES 1097 1097 5-hydroxylysine (Potential).
FT MOD_RES 1153 1153 3-hydroxyproline.
FT CARBOHYD 254 254 O-linked (Gal...) (By similarity).
FT CARBOHYD 1354 1354 N-linked (GlcNAc...) (By similarity).
FT CONFLICT 1187 1187 F -> L (in Ref. 5).
FT CONFLICT 1441 1441 Q -> H (in Ref. 6).
SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;
Query Match 57.7%; Score 2679; DB 1; Length 1453;
Best Local Similarity 53.7%; Pred. No. 7e-100;
Matches 540; Conservative 30; Mismatches 250; Indels 186; Gaps 16;
Qy 1 GPPGPGPTGLPGPPGGRGSGRPGFGADGVAGPKGPA-----GER 42
Db 174 GPRGLPGPPCAPQPGQFGPPGPGFPGASGNGRPGAGPPGKNGDGEAGKPGRQGR 233
Qy 43 GSTPGAGPKSGPGEAGRPGAGLPGAKGLTSGPSGPGDKTGPAGQD----- 93
Db 234 GPRPGQARGLPCTAGLPGMKHGRFSGLDGAKGQPGAPGKGPSPGNGAPGQMPR 293
Qy 94 -----GRPGPPGPPGARGO---AGVWGPPGPKGAAGE---PGKAGRGVGGPPGAVGPA 141
Db 294 GLPGRGRPPSPGAGAGNDGAPGAGPPGFTGPGAGPGFPGFAAGAKGETPGQARGSE 353
Qy 142 GKDGAGAGQPPGPGAPGAGRGEGQP-----AGSPFGQLPGPAGPPGSA 186
Db 354 GPQSGRGPSPGPGFAGAGPAGNPGADGPGAKGATGATGAGIAGAPGPGARGSPGQGS 413
Qy 187 GKPGEQVPGDLGAP-----GPSGPA-----GEPGTGLPGPP 219
Db 414 GARGPKNSGEPGAPGNKGTGAKGFGGAVGPPGPGAGEGKRGARPEGPPAGLPGPA 473
Qy 220 GERGGGSRGPPGADGVAGFGKPGAGRGSPGPGAPKSGPGEAGRPGAGLPGAKGLTSP 279
Db 474 GERGAPSRGPPGADGAGIAGFKGPGRGSGPAGVPGKSGPCEAGRPGAGLPGAKGLTSP 533
Qy 280 GSPGPDGKTGPPGAGDGRPPGPPGARGQAGVWGPPGPKGAAGEPKAGRGVGP 339

Db 534 GSPGPDGKTGPPGAGDGRGPPGAGCGAGVGMFGPKGAAGRGKRGGERGAGP 593

Qy 340 GAVGAGKDGAGAGAGPPGAGPAGERGEGGQAGGPGFQGLPGPAGPPGEGAGKRGEGV 399

Db 594 GAVGAGKDGAGAGAGPPGPTGPPAGERGEGGQAGGPGFQGLPGPAGPPGEGAGKRGEGV 653

Qy 400 GDLGAPGSGPAGE-----PGP-----TGLPGPPGEGGP 429

Db 654 GNAGAPGAGARGGEGFRCRGVQPPGPPGPRGANGAPGNDGAGDAGAGAGGNGPP 713

Qy 430 GSRGPPGADGAVGPKGAPGERSGPP-----AGPKSGPGEAGRP---GEA 471

Db 714 GLEGMFGGERGAGLPGKAGDGRDPPKAGDAGPKDGLRGLTGRTGPPGAGAGDGEA 773

Qy 472 GLPGAKGLTSGSPGPKGTGPPGPA-----GQDGRGPP 507

Db 774 GPPGAGTARGAGDGRGEGPPGAGPAGPPGADGQPGAKGTGDAGAKDAGPPGPA 833

Qy 508 GPPGARGAGVGMFGPKGAAGEGPKAGERGVPGPGAVGAGKDGAGAGQPPGAGPA 567

Db 834 GPTGAPGAGZVGCAPGPKAGSAGPPGATGPPGAGRGVPPGPGSGNIGLPGPPGPKGZ 893

Qy 568 GE---RQGGPAGSGFQGLPGPAGPCEAGKRGEGVPGDLGAPGSGRAGEGPTGLP 624

Db 894 GSKRGSTGAPRGEGPPGAPGPPGPKGSGADGPIGAPGTPGQGIAGRGVVGLP 953

Qy 625 GPPGRRGPPGPPGADGAGKPKGAPGERSGPPAGP---KGSFGGAGRPGEAGLPGAK 681

Db 954 GQRGPPGLPGPSGEPKQPSGASGERGPPGPMGPPGLAGPPGEGAGGAGGAP 1013

Qy 682 GLTSGSPGPDGKTGPPGAGQDGRGPPGPPGARGAGVGMFGPKGAAGEGPKAGER 741

Db 1014 GRDGAAGPKGRGETGPPGAPGAPGPPGPKGSGADGPIGAPGTPGQGIAGRGVVGLP 1073

Qy 742 GVPGPPGAVGAGKDG-----AGAQGPP-----GPAGPAGER 774

Db 1074 GPAGPQGRGDKGETGEGDGRMKHGRFSLGQPPGPPGAGPQSGSAGPAGPPGPP 1133

Qy 775 BEQGPAGSPGQGLPGPAGPGEAGKRGEGVPGDLGAPGSPGAG 820

Db 1134 CSAGAAGKDLGLGPGRTGPPGRTGEGVGVGPPGPPGPPGPPG 1179

RESULT 11

OS 093251 PRELIMINARY; PRT; 1445 AA.

AC C93251; 01-NOV-1998 (TReMBLrel. 08, Created)

DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)

DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)

DE Alpha 1 type I collagen.

GN Name=alpha 1 type I collagen;

OS Rana catesbeiana (Bull frog);

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae; Rana.

OX NCBI_TaxID=8400;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99294154; PubMed=10367734;

RA Asahina K., Uch R., Osara M., Yoshizato K.;

RT "Cell-type specific and thyroid hormone-dependent expression of genes

RT of alpha 1(I) and alpha 2(I) collagen in intestine during

RT amphibian metamorphosis";

RL Matrix Biol. 18:89-103(1999).

DR EMBL; AB015440; BAA29028.1; --

DR GO; GO:0005581; C:collagen; IEA.

DR GO; GO:0005737; C:cytoplasm; IEA.

DR GO; GO:0005201; P:extracellular matrix structural constituent; IEA.

DR GO; GO:0006817; P:phosphate transport; IEA.

DR InterPro; IPR008161; Clg helix.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR000885; Fib collagen_C.

DR InterPro; IPR001007; VWF_C.

DR Pfam; PF01410; COLFI; 1.

DR Pfam; PF01391; Collagen; 18.

DR ProDom; PD000007; Clg_helix; 4.

DR ProDom; PD002078; Fib_collagen_C; 1.

DR SMART; SM00038; COLFI; 1.

DR SMART; SM00214; VWC; 1.

DR PROSITE; PS01208; VWF_C_1; 1.

DR PROSITE; PS01184; VWF_C_2; 1.

DR Collagen.

Qy 1445 AA; 137251 MW; F59BB550C9873F04 CRC64;

Query Match 55.3%; Score 2568; DB 2; Length 1445;

Best Local Similarity 53.6%; Pred. No. 1.9e-95;

Matches 516; Conservative 43; Mismatches 259; Indels 144; Gaps 17;

Qy 1 GPPGEPGPTGLPQPPGERGPGS-----RGFPGADGVAGPKGPA-----GER 42

Db 221 GXPGRPPGPPGPPGQARGLPGTAGLPGMKHGRFNGLDGAKGDTGPAGPKGEPGNPGEN 280

Qy 43 GSPGAPGKSGPCEAGRPCEAGLPKAGLGTGSPGSPGDKTGPPGP-----A 90

Db 281 GAFQGVPPGLFGERGPPGSPGAGARGNDGTPGAAGPPGTGTPGPPGPGVGPCKGDA 340

Qy 91 GQDGRPPPPGAGAGQAGVGMFGPKGAAGBPAGKAGERGVPGPGVAGVPAGKDGAGAG 150

Db 341 GFGGRRGPPGPPGAGGPPGAGQAGPAGSAGNPDTGQPGAKGATGAPGAGPFPGAR 400

Qy 151 GPPGAPGAGERGEGGPPGSPGFPQGLPGPAGPPGAGKPGEGVPGDLGAPGSPGAPGE 210

Db 401 GAFGQPPGSGSPGKNNNGEPGAKNGEPGKAGSGPAGSQGPPGPPGEGGKSGRGP 460

Qy 211 GPTGLPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 270

Db 461 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 520

Qy 271 GAKGLTSGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 330

Db 521 GAKGLTSGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 580

Qy 331 GERGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 390

Db 581 GERGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 640

Qy 391 GXPGEQGVPGD-----LGAPGSGP-----AGEP 414

Db 641 GXPGEQGVPGDVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 700

Qy 415 -----GPTGLPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 450

Db 701 GAPGGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 760

Qy 451 GSP-----GPAGP---KGSPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 498

Db 761 GAPGPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 820

Qy 499 GQDGRPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 558

Db 821 GDAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 880

Qy 559 GPP 615

Db 881 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 940

Qy 616 GEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 675

Db 941 GTRGVTPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1000

Qy 676 GLPGAGL-----TGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 714

Db 1001 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1060

Qy 715 GARGAGVGMFGPPGPKGAGGPKKAGERG-----VGPBPAGVAGPKDGAGAG 762

```
Db 1061 GPAGPAGARGSGAGARGDKGAGEQGERGMKCHRGFNDLPDPGAPGAGHAGEGPGSGAS 1120
QY 763 GPPGPAGPAGRGQGPAGSGFQGLPFPAGPPGAGKPGSE---QGVPGDLGAPGPGGA 819
Db 1121 GPAGPRGPGSGSGPGKDGSG---NGLPGIPGPPGRKRTGDVGAPGPGPAGPPGPPGP 1177
QY 820 GG 821
Db 1178 GG 1179

RESULT 12
QYIB4
ID Q9YIB4 PRELIMINARY; PRT; 1450 AA.
AC Q9YIB4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen.
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX NCBI_TaxID=8330;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Regenerate forelimbs;
RX MEDLINE=99407244; PubMed=10474166;
RA Asahina K., Obara M., Yoshizato K.;
RT "Expression of genes of type I and type II collagen in the formation
RL and development of the blastema of regenerating newt limb.";
RL Dev. Dyn. 216:59-71(1999).
DR EMBL; AB015438; BAA36973.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; E:extracellular matrix structural constituent; IEA.
DR GO; GO:0006811; P:phosphate transport; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib_collagen_C.
DR InterPro; IPR001007; WFC_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; C_helix; 4.
DR ProDom; PD02078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; WVC; 1.
DR PROSITE; PS01208; WFC1; UNKNOWN_1.
DR PROSITE; PS0184; WFC2; 1.
KW Collagen.
SQ SEQUENCE 1450 AA; 137563 MW; ABF8A74841B87B7C CRC64;

Query Match 55.2%; Score 2561.5; DB 2; Length 1450;
Best Local Similarity 52.3%; Pred. No. 3.4e-95;
Matches 517; Conservative 39; Mismatches 261; Indels 171; Gaps 17;

QY 1 GPPGP-----GPTGLGPPGPERG-----GP----- 21
Db 192 GEFPGAAGALGRLGLPFPKNGDDGESKGRPRGPRGSGPGQAGRLPFTAGLPWK 251
QY 22 GSRGFPAGDVAGKPGA-----GERSPGAPKPGSPGAGRPGEAGLPGAKGLT 72
Db 252 GHRGNGLDKAGDNGPAGKPEPGNPGENGAPGQAGPRGLFGERGRGAGPGAGRND 311
QY 73 GSPGSPGPGKTPGPP-----AGDGRPPGPPGARGQAGVNGFPGPGKAA 120
Db 312 GSPGAAGPPGTPGTPGPPGAVGAKGDAGPQSGRSGEGPQARGEPGA---PGPAGAA 368
QY 121 GEPKXAGEGVPGPAGVP---AGKDGAGAQGPAGPAGGERGEGPAGSGFQGLP 177
Db 369 GSGNPPTDQGGKGAATGSPGIAGAPFPFGARGAPPGQAGAPGPKNGNGEPQAQNK 428
QY 178 GPAGPPGAGKFGEGQVFDGLGAPGSPGAPGPTGLPFPGERGPGSGRGPAGDGA 237
```

```
Db 429' GEPGAKGERGPAVQVPPGPSGEEKRGSRGEPGPPGAPGAGERGGPSGRGPGSGDAS 488
QY 238 GKXGPAGERGSPGAPGKSPGEAGRPGHAGLPGAKLGTSGSGSPGDPKTKTGPAGQD 297
Db 489 GPKGAPGERGSGVAPGKSTGESRPGEPGLPGAKLGTSGSGSPGDPKTKTGPAGAAGD 548
QY 298 GRPGRGPPGARGQAGVNGFPGPKGAAGEPGKAGRGVPGPAGVGPAGKDGEGAGAQGP 357
Db 549 GHPPGPPGARGQAGVNGFPGPKGAAGEPGKAGRGVPGPAGVGPAGKDGEGAGAQGP 608
QY 358 GPAGPAGERGEGPAGSPGFGLPAGPAGPAGKPGEGVPGDLGAPGSPGAGE--- 413
Db 609 GPSGSPGERGEGPAGSPGFGLPAGPAGPAGKPGEGVPGDLGAPGSPGAGE--- 668
QY 414 -----PGPTGLGPP-----PGERGGPGSGRGP----- 435
Db 669 GERGGGPAGAGPGRGSPGNDGAKGAGAGAPGGRGPPFOLQMPGERSAGMPGAK 728
QY 436 -----GADGVAGPKGPPAGRGSP-----GPAGPKSGPGEAGRP 468
Db 729 GDRGDAGTKGADGAPGKDGARGLTGPFGPPGSGAPGDKGEGSPGAPGTGARGSPGR 788
QY 469 GEAGLPGAKLGTSGPS-----PGPDGKTGPPGAGQDGRPGPPGPGKARQAGVMGFP 525
Db 789 GEPGAPCPAGICGPPGADGQPGKAGSGDAGPKGADGAPGAGTGAAPGAGNVGAPGP 848
QY 526 GAAGEPKKAGERGVPGPAGVGPAGKDGAGAGQGGPPGAPGAGERGEGPAGSPGFG 585
Db 849 GTRGAGPAGTATGFAAGELGPPGPGNAGPPGPGPGKSGKSGRGTGPAGRSBP 908
QY 586 GPAGPP---GEAKPGEGVQVGDLAGPDSGSPAGEPGPTGLPQPPGERGPGSRGPPGAD 642
Db 909 GPAGPPGSPGKSGPSGSDGAPAGIPGPGIAGRGVVGLPQGRGERGFSGLPGPAGP 968
QY 643 GVAGPKPAGERGSPGAPG-----KGSPEAGRPGEAGLPGAK-----GLT 684
Db 969 GKQPSGPNGERGPPGPPGPGGPGGREGSGSGAGRGDSRPGKDRGNGPS 1028
QY 685 GSPGSPGDKTGPAGQD---GRPPGPPGARGQAGVNGFPGPKAAGEPGKAGE- 740
Db 1029 GPPGAPGAPGAPGVGPAKNGDRGTGPAGPAGPAGPSGVRGAPGAPGARGDKGAGEQ 1088
QY 741 -----RGVPGPAGVGPAGKDGAGAGQGGPPRAGPAGERGEGPAGSPGFG 792
Db 1089 GERGMKGRGFGNMGQPPPGSSGEGQAGPGSGPAGPRGPGSGSGTGDVNGLP 1148
QY 793 GPPGKAGKPGEGVQVGDLAGPSPGAPG 820
Db 1149 GPFGRGRNDVGPAGPPGPPGPPG 1176
```

RESULT 13

```
Q802B5
ID Q802B5 PRELIMINARY; PRT; 1449 AA.
AC Q802B5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cellial-prov protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932;
RA Srausberg R.L., Reingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
```

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny K.D., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzyzanski M.T., Skalska U., Smalios D.E., Schmerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP
 RC TISSUE=Embryo;
 RX MEDLINE=22341132; PubMed=12454917;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT initiative";
 RL Dev. Dyn. 225:384-391(2002).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP
 RC TISSUE=Embryo;
 RA Klein S., Strausberg R.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC049829; AAH49829.1; -;
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLF1; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; VWC; 1.
 DR ProDom; PD000007; Clg helix; 4.
 DR ProDom; PD002078; Fib collagen_C; 1.
 DR SMART; SM00038; COLF1; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWF_C_1; UNKNOWN_1.
 DR PROSITE; PS0184; VWF_C_2; 1.
 KW Collagen.
 SQ SEQUENCE 1449 AA; 137464 MW; 4B32E20BECCE978C CRC64;

Query Match 54.5%; Score 2529; DB 2; Length 1449;
 Best Local Similarity 54.6%; Pred. No. 57e-94;
 Matches 500; Conservative 53; Mismatches 261; Indels 102; Gaps 13;

QY 1 GPPGPGTGLPGRGGRGPGS-----RGFGADGVAGPKGPA-----GER 42
 Db 222 GKPRGPRGRGPPGQAGRLPCTAGLPGMKHGRGNGLDGAAGSDGAPGKPGFPGSPEN 281
 QY 43 GSPGAPGKPGFAGRGEGALPGAKLTSGPSGPGDKTGPDP-----A 90
 Db 282 GAPGVGRGSLGSRGPRGPGSGARGNDGAPGAAGPSTGSPGPPGPGVGPKGDA 341
 QY 91 GQDGRGPPGPGARGQAGVGMFPKPGAAGEPGKAGRGVPPGPAAGVAGPKDGAAG 150
 Db 342 GPGSGSDGPGQGRGEGPAPGQAGAGSPGSDGQPKAGKATGATGATGAPGPFGR 401
 QY 151 GPPGAPGAGRGPGGPGAGSPGFGPLPGPAGPPEAGKPGEGVPGDLGAPGSGPAGEP 210
 Db 402 GAPAGQGGSGPGKGNNGEPGAGKNGEAGKAGEPGAGVQGPVPPGSGEGKRGSGEP 461
 QY 211 GPTGLPGRGPGSGRFGGADGVAGPKGAPGAGRGSPGAPGKSPGAGRPGAGLP 270
 Db 462 GFAGPPGAPGAGRGSGRFGGSDGAGPKGAPGGERGVGAPGPKSGSGSGRFGPGLP 521

QY 271 GAKGLTSGPSGPGDKTGPAGPAGQDGRPRGPPGARGQAGVMGFPGPKGAAGEPGKA 330
 Db 522 GAKGLTSGPSGSDGKTGPAGAPQDGRAGPPGARGQAGVMGFPGPKGAAGEPGKN 581
 QY 331 GERGVPPGAVGVPAGKAGEAGAAQPPGPPGAPGARGGQGPAGSPGFOGLPGPAGPPGEA 390
 Db 582 GERGVAGPPGAVGLPGKODAGAQGPAGPPGAPGARGGQGPAGSPGFOGLPGPAGPPGES 641
 QY 391 GXEGEGVPGDLGARGSPGAGE---PQPTGLPGRGGRGPGSGRFGGADGVAGPKGA 447
 Db 642 GRPGEVGVDPDVPSPGAPGARGRGFPGERGAQGPQGARGSGAPENDGAKGEAGAA 701
 QY 448 GERSGPPGAPGKSGPGEAGRPGEAGLPKAK-----GLTSGPSGPGDKTGPAGD 501
 Db 702 GAPGGQPPGLQGMFGER---GSSGLPGAKGDRGQGVKGSDDTGPDKGVRLTGPGR 757
 QY 502 GRPGRPPGARGQAGVMGFPGPKGAAGEPGKAGRGVPPGPAAGVAGPKDGEAGAA 558
 Db 758 --PGGAPGDKGEGAPGAPGAGTGAAGPGERGSGPPGAPGAGPPGADGQPKAGEQ 815
 QY 559 -----GPPGPAGPA-----GERGQGPAGSPGFOGL-----PGP- 587
 Db 815 GDSGAKGDAGPFPAGTGAPOFAGALGSPGPKGARGAPPGATGTFGAAGRLGPPGPS 875
 QY 588 --AGPPGEGAKFGEQVPGDLGAPGSPGAPGPTGLPGRGGRGSGSGRFGPGADGVA 645
 Db 876 GNAGPPGSPGAPGKAGKPRGTGPRGSGEPGAAGPPGPPGKSGSGSDGAPAGP 935
 QY 645 GPKGAPGERSGPPGAPGKSGPGEAGRPGEAGLPKAKGLTSGPSGPGDKTGP 702
 Db 936 GFGVAGSRGTGLPGRGGRGFGSLPGPAGEPGKQGSFGSGRGGPPGSGPGLGPP 995
 QY 703 GQDGRPPGPPGARGQAGVMGFPGPKGAAGEPGKAGRGVPPGPAAGVAGPKDGEAGAA 762
 Db 996 GSGREGAPGSECAPGRDCAVGPKDRGEGGAPGAPGAPGAPGAPGVPYGPAGKSGDRGT 1055
 QY 763 GPPGPAGPA-----GERGQGPAGSPGFOGL-----PGP- 804
 Db 1056 GSPGAPGAGTAGAGAPGAPGPRGDKGAGEGQGRGKGRGNGSPSGPPGPPGSGEQ 1115
 QY 805 GVEGDLGAPGSPGAP 820
 Db 1116 GPSGASGAPGPPG 1131

RESULT 14
 Q6PCL3
 ID Q6PCL3 PRELIMINARY; PRT; 1225 AA.
 AC Q6PCL3;
 DT 05-JUL-2004 (TREMELrel. 27, Created)
 DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)
 DE Colial protein.
 GN Name=Colial;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mix FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Hopkins R.P., Zeeberg B., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Altschuld S.F., Jordan H., Moore T., Farmer A.A., Rubin G.M., Hong L.,
 RA Diatchenko L., Marusina K., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

```

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Mix FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RC Strausberg R.; to the EMBL/GenBank/DBJ databases.
RL Submitted (Oct-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC059281; AAH59281.1;
DR InterPro; IPR008161; Cig_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR009041; PNP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLF1; 1.
DR Pfam; PF01391; Collagen; 14.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Cig_helix; 2.
DR ProDom; PD02078; Fib_collagen_C; 1.
DR SMART; SM00038; COLF1; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS0184; VWFC_2; 1.
DR KW Collagen.
SQ SEQUENCE 1225 AA; 117860 MW; B6886CBB4457F4D9 CRC64;

Query Match 53.8%; Score 2494; DB 2; Length 1225;
Best Local Similarity 56.0%; Pred. No. 1.5e-92;
Matches 510; Conservative 29; Mismatches 227; Indels 144; Gaps 20;

QY 12 POPPERGG-----PGRGPGCAGVAGPKGAGERSGPGAGPKGSPGGEAGRGGLP 66
DB 73 PNFQRECECCAFCEEYVNSDEV-VGEGPKGDPGPGQPRGPGVPPGPRGDI PQGGLP 131
QY 67 GAKGLTSGPSGPPDGK-----TGPPGAGQDGRGPGPGARGQA 108
DB 132 GPPGHGPPGPPGLGNGFASQMSYGYDEKSAVSVPGMSPGSRGLPGPGAGPQ--- 188
QY 109 GVMGPPGKGAAGBPGKAGERGVPGPGAVGPGAGKDGAGAGAGGPP---GPAGPAGERGEQ 165
DB 189 ---GFGPPGPEBEGCGSGPMGPRGPPGPGKXGNDGAGKFGPRGERGPGPGQARGLP 245
QY 166 GPAGSP-----GFQGLPAGPPGEGAKPGEGQGVPGDLGAPSGSPGAGEGPTGLPGPP 219
DB 246 GTAGLPGMKHGRGFSGLDCAKGDAGPAGPKGEPGSGFENGAFQMGPRGLPCRGRFGPP 305
QY 220 GSRGPGSGRPGCAGVAGPKGAGERSGPGAGPKGSPGAGRGPGAGLPCAAGLGTSP 279
DB 306 GT---AGARGDGAAGVAGPPT-----GTPGPPGPGVAGKAGEAGPQAR----- 350
QY 280 GSPGPDGKTGPPGPAQDGRPPGPPGARGAGVGMFPFGKGAAGEPKGAGERGVPGPP 339
DB 351 GSEGGQGVTRGEPGPPGPAAGPAGNPGADGQ-----FGAKGANGAPGIA---GAPGPP 401
QY 340 GAVGPAKDGEGAGAQPPGPPGAGPAGERGEGQAGSGFGQLGPGPAGPGEAKPGEQVPP 399
DB 402 GARGP-----SGPQPSGPPGPKNSGE-----PGAPGNKGGDTGAKGEPGATGVQGPP 449
QY 400 GDLGAPGSPGAGEPPTGLPGPBERGGPGSRGFPFGADGVAGPKGAGERSGPGAGPK 459
DB 450 GPAGEGKKGARGEPGSLPFPFGGERGPGSRGFPFGADGVAGPKGSGERGAPGAPGPK 509
QY 460 GSPGEAGRPGEAGLPGAKGLTSGPSGPPDGKTGTPGPPAGQDGRGPPGPPGARGQAGVM 519

```

```

DB 510 GSPGEAGRPGEAGLPGAKGLTSGPSGPPDGKTGTPGPPAGQDGRGPPGPPGARGQAGVM 569
QY 520 GPPGKGAAGBPGKAGERGVPGPGAVGPGAGKDGAGAGAQGPPGPGPAGERGEGQAGSP 579
DB 570 GPPGPKGTAGBPGKAGERGLGPPGAVGPGACDGEAGAGQAGPAGPAGERGEGQAGSP 629
QY 580 GFQGLPAGPAGPGEAGKPGEGQGVPGCDLGAPOSPGAGE-----PGP----- 620
DB 630 GFQGLPAGPAGPGEAGKPGEGQGVPGCDLGAPOSPGAGE-----PGP----- 689
QY 621 -----TGLPGPGERGGPSRGFPFGADGVAGPKGAGERSGPGPAGPKGSPGE- 668
DB 690 GAPGNDGAKGDTGAPGAPGAGSQGAPLQGMPPGERGAGLPGPKGDEGDAGPKGADGSPGKD 749
QY 669 -----AGRPGEAGLPCAAGLTCSPSPGPDGKTGTPGPPAGQDGRGPPGPPGARGQA 720
DB 750 GARGLTGTPGPPGPAAGPDKGEGAGSPGPGTARGAGPGRGEGAGPFPAGFAGPPGAP 809
QY 721 GVMGPPGKGAAGBPGKAGERGVPGPGAVGPGAGKDGAGAGAQGPPGPGPAGERGE--- 776
DB 810 GAGGAPGVPVGPAGKNGDRGEGTGPAGPAGPIGAGARGPAGPGQGRDKGEGQDGRGK 869
QY 777 -----QGPAGSPGFQGLPQPA-----GPPGEGAGKPGE---QVPGDLGAPGP- 815
DB 870 GHRGFSGLQGPFGSPGSGEQGSGASGAPGPRGPGSAGSPCKDGLNGLPGPIGPPGPR 929
QY 816 -----SGPAG 820
DB 930 GRTGDSGPAG 939

RESULT 15
ID AAH59281 PRELIMINARY; PRT; 1225 AA.
AC AAH59281;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE Collal protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary tumor;
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary tumor;
RA Strausberg R.;
RA Submitted (Oct-2003) to the EMBL/GenBank/DBJ databases.
RL

```

DR EMBL: BC059281; AAH59281.1; --
SQ SEQUENCE 1225 AA; 117860 MW; B6B86CBB457F4D9 CRC64;

Query Match 53.8%; Score 2494; DB 2; Length 1225;
Best Local Similarity 56.0%; Pred. No. 1.5e-92;
Matches 510; Conservative 29; Mismatches 227; Indels 144; Gaps 20;

```
QY 12 PCPPGERGG-----PGRGFGGADGAGVAGPKGAGRGSGPGPAGPKGSGPGEAGRGFGEGACLP 66
Db 73 PNPQRREGCCAFCEEEYVSPNSDEV-GVEGPKGDPGPGPRGPFVGPFGDRGDFGQFGLP 131
QY 67 GAKGLTSGSPGSPGDDCK-----TGPPGPGADGDRGPPGPPGARGQA 108
Db 132 GPPGHGPPGPPGLGNGFASQMSYCYDEKSAGSVFPGMGSGRGLPGPFCAFPQ--- 188
QY 109 GVMGFPKGAAGEPKAGRGVPGPPGAVGAPGAKDGAAGAGGPP---GPAPAGERGEQ 165
Db 189 ---GFQGGPGEFEGSGGPMGPRGPPGPGKNGDDGEAGKGRFGRGPPGPGQARGLP 245
QY 166 GPAGSP-----GFOGLPGAPGPGGAGKPGGQVGGDLGAPGSPGAGEPGTGLPGPP 219
Db 246 GTAGLPGMKHGRFSELGAKGDAGPAGPKGEPGSPGNGAPGQMGPRGLPGERGRGPP 305
QY 220 GERGGSGRGPFGADGAVGPKGAGERGSPGAPGKSGPGEAGRPGEAGLPGAKGLTGP 279
Db 306 GT---AGARGNDGAVGAAGPPGPT-----GPTGPPGPPGAVGAKGEAGPQGAR----- 350
QY 280 GSPGPDGKTGPPGAGQDGRGPPGPGARGOAGVMGFPKGAAGBPKAGRGCVDPGP 339
Db 351 GSEGGQVGRGPPGPPGAGAGPAGNPGADGQ-----PGAKGANGAPGIA---GAPGFP 401
QY 340 GAVGPAGKDGAGAGQGGPPGAPGAGERGEGGAGSPGFGGLPGPAGPPGAGKPGEGQVP 399
Db 402 GARGP-----SGPQGPSPGPKGNCE-----PGAPGNKGDGTGAKBPGATGVQGP 449
QY 400 GDLGAPSGPAGRPGTGLPGPGERGPGSRGFPDAGVAGPKGAGRGSGPAGPK 459
Db 450 GPAGEEGRGARGPFGPSGLPGPGERGGSGRFPGADGAVGPKGSGRGAPGAPGPK 509
QY 460 GSPGAGRPAGLPGAKGLTSGSPGPDGKTGPPGAGDGRGPPGPPGARGGAGVM 519
Db 510 GSPGAGRPAGLPGAKGLTSGSPGPDGKTGPPGAGDGRGPPGPPGARGGAGVM 569
QY 520 GPPGPKGAGBPKGAGRGVPGPPGAVGPKDGEAGAGQGGPPGAPGAGRGGQGPAGSP 579
Db 570 GPPGPKTAGBPKGAGRGVPGPPGAVGPKDGEAGAGQGGPPGAPGAGRGGQGPAGSP 629
QY 580 GFQGLPGAPGPPGAGKPGEGVPGDLGAPGSPGAGE-----PGP----- 620
Db 630 GFQGLPGAPGPPGAGKPGEGVPGDLGAPGSPGAGE-----PGP----- 620
QY 621 -----TGLPGRGPGSGRGPFCADGAVGPKGAPAGERGSPGAPGKSPGE- 668
Db 690 GAPNGDAKAGDTGAPGAPGAGPGLQMPGERGAAGLPGKDRGADGPKGADGSPGXD 749
QY 669 -----AGPGEAGLPGAKGLTSGSPGPDGKTGPPGAGDGRGPPGPPGARGQA 720
Db 750 GARGLTGPIGPPGAPGAPGKAGRGSPGPGTGAAGPDRGAGPPGPPGAPGAGPPGAP 809
QY 721 GVMGFPKGAAGBPKGAGRGVPGPPGAVGPKDGEAGAGQGGPPGAPGAGRGGE--- 776
Db 810 GAPGAPGVGAPGKNGDRGTGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 869
QY 777 -----QGPASGPGQGLPGPA-----GPPGEAGKPGE---QGVPGDLGAPGP- 815
Db 870 GHRGFGGLQGPSPGSGPGEAGSAGPAGRPFGSAGSPGKGLNGLPGIPGPPGR 929
QY 816 -----SGPAG 820
Db 930 GRTGDSGPAG 939
```